OM of: US-09-303-518D-653 to: PIR_71:* out_format : pfs

Date: Jun 30, 2002 8:13 AM

About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:
-MODEL=frame+_n2p.model -DEV=xlh
-Q=/cgn2_1/USPTO_spool/US09303518/runat_28062002_142713_4317/app_query.fasta_1.23501
-DB=PIR_71 -QFMT=fastan -SUFFIX=rpr -GAPOP=12.000 -GAPEXT=4.000
-MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000 -QGAPOP=4.500
-QGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500 -FGAPOP=6.000
-FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500 -DELOP=6.000
-DELEXT=7.000 -START=1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=100 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0
-ALIGN=45 -MODE=LOCAL -OUTFMT=pfs -NORM=ext -HEAPSIZE=500
-MINLEN=0 -MAXLEN=2000000000 -USER=US09303518_@CGN1_1_1092
-NCPU=6 -ICPU=3 -LONGLOG -DEV_TIMEOUT=120 -WARN_TIMEOUT=30
-NO_XLPXY -WAIT -THREADS=1

Search information block:
Query: US-09-303-518D-653
Query length: 4407
Database: PIR_71:*
Database sequences: 283138
Database length: 96089334
Search time (sec): 504.980000

score_list:

Sequence	Strd	Orig	ZScore	EScore	Len	! Documentation ..
pir2:C64057	+	2494.00	2502.49	3.7e-132	709	! IgA-specific metalloendopeptida
pir2:A37023	+	1614.00	1610.05	8.8e-83	1541	! IgA-specific metalloendopeptic
pir2:S61314	+	1605.00	1600.88	2.8e-82	1561	! IgA-specific metalloendopeptic
pir2:H64106	+	1587.50	1582.52	2.7e-81	1694	! IgA-specific metalloendopeptic
pir2:B41859	+	1587.00	1582.87	2.9e-81	1545	! IgA-specific metalloendopeptic
pir2:A81018	+	1585.50	1582.07	3.4e-81	1431	! serine-type peptidase NMB1998
pir2:A41859	+	1583.50	1578.45	4.6e-81	1702	! IgA-specific metalloendopeptic
pir2:A26039	+	1565.50	1561.32	4.6e-80	1532	! IgA-specific metalloendopeptic
pir2:C41859	+	1515.50	1509.28	3.0e-77	1849	! IgA-specific metalloendopeptic
pir2:A81937	+	1515.00	1509.17	3.2e-77	1773	! IgA-specific metalloendopeptic
pir2:C81169	+	1486.00	1479.78	1.4e-75	1815	! IgA-specific metalloendopeptic
pir2:I54632	+	1125.50	1119.67	2.0e-55	1377	! tsh protein - Escherichia coli
pir2:S57664	+	996.00	989.47	3.7e-48	1366	! IgA-specific metalloendopeptic
pir2:S61315	+	883.50	885.64	6.1e-42	496	! IgA-specific metalloendopeptida
pir2:S61319	+	876.50	878.38	1.5e-41	508	! IgA-specific metalloendopeptida
pir2:S61328	+	875.50	877.58	1.7e-41	497	! IgA-specific metalloendopeptida
pir2:S61323	+	868.00	870.05	4.5e-41	496	! IgA-specific metalloendopeptida
pir2:S61325	+	860.00	862.00	1.3e-40	496	! IgA-specific metalloendopeptida
pir2:S61329	+	859.50	861.28	1.4e-40	508	! IgA-specific metalloendopeptida
pir2:S61322	+	858.50	860.49	1.5e-40	496	! IgA-specific metalloendopeptida
pir2:S61327	+	856.50	858.48	2.0e-40	496	! IgA-specific metalloendopeptida
pir2:S61320	+	851.50	853.23	3.8e-40	508	! IgA-specific metalloendopeptida
pir2:S61330	+	845.50	847.20	8.3e-40	508	! IgA-specific metalloendopeptida
pir2:S61331	+	840.00	841.68	1.7e-39	507	! IgA-specific metalloendopeptida
pir2:S61332	+	718.50	720.31	1.1e-32	462	! IgA-specific metalloendopeptida
pir2:S61333	+	695.50	697.00	2.1e-31	471	! IgA-specific metalloendopeptida
pir2:T00317	+	629.50	621.23	1.3e-27	1300	! probable serine proteinase esp
pir2:B41500	+	407.00	408.65	2.9e-15	384	! IgA-specific metalloendopeptida
pir2:A65044	+	359.00	347.37	1.9e-12	1569	! hypothetical protein b2647 - E
pir2:B90674	+	348.00	337.86	7.5e-12	1327	! AidA-I adhesin-like protein [i
pir2:E85524	+	348.00	337.70	7.5e-12	1349	! probable beta-barrel outes mem
pir2:C91068	+	345.50	333.78	1.1e-11	1571	! hypothetical protein ECs3515 [
pir2:D85912	+	344.50	333.03	1.2e-11	1528	! hypothetical protein ypjA [imp
pir2:E95206	+	328.50	306.42	1.2e-10	4776	! cell wall surface anchor famil
pir2:T08658	+	295.50	288.14	6.1e-09	949	! hypothetical protein f949 - Esc
pir2:G64993	+	292.50	282.57	9.5e-09	1250	! yfaL protein - Escherichia col
pir2:D91018	+	290.50	280.56	1.2e-08	1250	! hypothetical protein ECs3116 [
pir2:F85862	+	290.50	280.56	1.2e-08	1250	! hypothetical protein yfaL [imp
pir2:B43855	+	279.50	267.96	5.2e-08	1477	! high-molecular-weight surface-
pir2:A35140	+	277.50	265.34	6.9e-08	1577	! hemolysin A precursor - Proteu
pir2:S15204	+	275.50	268.28	8.0e-08	922	! pertactin - Bordetella parap
pir2:A47675	+	274.00	266.88	9.7e-08	911	! 68K outer membrane protein P
pir2:F90073	+	274.00	258.45	1.2e-07	2271	! hypothetical protein SA2447
pir2:C82199	+	273.50	251.52	1.4e-07	4558	! RTX toxin RtxA VC1451 [impc
pir2:B89921	+	273.00	247.44	1.6e-07	6713	! hypothetical protein ebhA [
pir2:F81045	+	271.00	254.50	1.7e-07	2514	! hemagglutinin/hemolysin-rel
pir2:AF0351	+	270.50	259.20	1.7e-07	1430	! probable autotransporter pr
pir2:A32560	+	268.00	260.86	2.1e-07	910	! outer membrane protein P.69
pir2:JN0896	+	267.50	254.89	2.5e-07	1645	! crystalline surface layer p
pir2:A43855	+	267.00	255.02	2.6e-07	1536	! high-molecular-weight surfa
pir2:A41477	+	261.50	245.97	5.8e-07	2249	! 190K surface antigen precur
pir2:AI0484	+	258.00	249.30	7.9e-07	1070	! probable autotransporter pr
pir2:AF0394	+	258.00	243.96	8.8e-07	1910	! probable adhesin hmwA [impc
pir2:T09083	+	254.50	238.83	1.4e-06	2273	! hemagglutinin/hemolysin-rel
pir2:AB3528	+	254.50	237.75	1.5e-06	2554	! extracellular serine protei
pir2:D71917	+	254.00	235.18	1.6e-06	3194	! toxin-like outer membrane p
pir2:B81192	+	253.00	238.62	1.7e-06	1975	! hemagglutinin/hemolysin-rel
pir2:A97859	+	250.50	235.89	2.3e-06	2021	! 190-KDa cell surface antige
pir2:C48399	+	250.00	235.39	2.5e-06	2020	! ABC-type transport protein
pir2:H71879	+	248.50	232.29	3.1e-06	2399	! toxin-like outer membrane p
pir2:S28634	+	247.50	237.04	3.2e-06	1286	! adhesin AIDA-I precursor -
pir2:A98014	+	247.00	240.52	3.1e-06	836	! hypothetical protein ECs3081
pir2:C85858	+	247.00	240.52	3.1e-06	836	! hypothetical protein yejO [i
pir2:AD0123	+	245.50	225.26	5.0e-06	3705	! probable autotransporter pr
pir2:AE3259	+	245.00	233.92	4.4e-06	1374	! extracellular serine protei
pir2:G64964	+	244.00	235.04	4.8e-06	1091	! hypothetical protein b2000
pir2:T02345	+	243.50	229.96	5.6e-06	1791	! hypothetical protein KIAA03
pir2:G81044	+	243.50	228.97	5.8e-06	1995	! hemagglutinin/hemolysin-rel
pir2:D64988	+	242.50	235.99	5.6e-06	836	! yejO protein - Escherichia c
pir2:A64596	+	241.50	231.36	6.8e-06	1238	! hypothetical protein HP0609
pir2:E83641	+	241.50	221.67	8.3e-06	3535	! probable hemagglutinin PA0C
pir2:T14652	+	241.00	229.13	7.5e-06	1492	! protein J - Yersinia pestis
pir2:T14966	+	241.00	228.81	7.6e-06	1545	! phage lambda-related host s
pir2:JQ0405	+	240.50	231.39	7.6e-06	1106	! hypothetical 119.5K protein
pir2:T31105	+	240.00	217.12	1.1e-05	4919	! hypothetical protein 2 - Ha
pir2:C83339	+	240.00	215.87	1.1e-05	5627	! hypothetical protein PA2462
pir2:F83068	+	239.00	223.73	1.0e-05	2154	! hypothetical protein PA4625
pir2:H98323	+	238.50	227.60	1.0e-05	1341	! hypothetical protein AGR_L_
pir2:D90011	+	238.50	221.92	1.1e-05	2481	! FmtB protein [imported] - S
pir2:D81411	+	238.00	230.01	1.0e-05	978	! probable lipoprotein Cj0629
pir2:H97912	+	237.50	224.05	1.2e-05	1767	! conserved hypothetical prot
pir2:D71630	+	237.00	224.22	1.3e-05	1643	! outer membrane protein B pr
pir2:JC1340	+	237.00	224.17	1.3e-05	1651	! outer membrane protein B pr
pir2:E97835	+	237.00	224.15	1.3e-05	1655	! hypothetical protein rompB
pir2:A64905	+	235.00	224.19	1.6e-05	1325	! ydeK protein - Escherichia
pir2:T34513	+	235.00	215.21	1.9e-05	3507	! hypothetical protein ZK783.
pir2:AH2515	+	234.50	211.55	2.2e-05	4936	! hypothetical protein alr730
pir2:AB2018	+	233.50	218.89	2.1e-05	1999	! hypothetical protein all169
pir2:S40431	+	233.00	233.96	1.6e-05	370	! IgA-specific metalloendopept
pir2:G84693	+	233.00	225.84	1.9e-05	891	! probable proline-rich protei
pir2:B81989	+	232.50	217.81	2.4e-05	2015	! hypothetical protein NMA068
pir2:AF2959	+	232.00	223.30	2.3e-05	1052	! conserved hypothetical prot
pir2:H81193	+	230.00	212.58	3.5e-05	2703	! hemagglutinin/hemolysin-rel
pir2:G64866	+	229.50	221.68	3.1e-05	955	! probable membrane protein b1
pir2:AB0204	+	229.50	219.07	3.2e-05	1268	! conserved hypothetical prot
pir2:B72078	+	229.50	218.11	3.3e-05	1407	! polymorphic outer membrane
pir2:B99789	+	229.00	218.56	3.4e-05	1268	! hemagglutinin/hemolysin-rel

seq_name: pir2:C64057

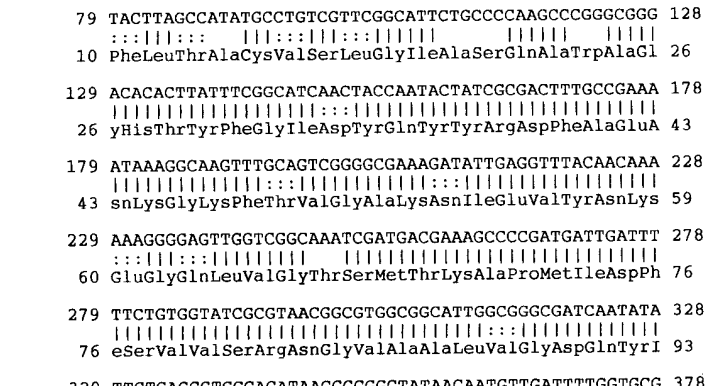
seq_documentation_block:
IgA-specific metalloendopeptidase type 1 homolog - Haemophilus influenzae (strain Rd
C;Species: Haemophilus influenzae
C;Date: 18-Aug-1995 #sequence_revision 18-Aug-1995 #text_change 31-Oct-1997
C;Accession: C64057
R;Fleischmann, R.D.; Adams, M.D.; White, O.; Clayton, R.A.; Kirkness, E.F.; Kerlavage
; Gocayne, J.D.; Scott, J.; Shirley, R.; Liu, L.I.; Glodek, A.; Kelley, J.M.; Weidman
, D.M.; Brandon, R.C.; Fine, L.D.; Fritchman, J.L.; Fuhrmann, J.L.; Geoghagen, N.S.M.
Science 269, 496-512, 1995
A;Authors: Gnehm, C.L.; McDonald, L.A.; Small, K.V.; Fraser, C.M.; Smith, H.O.; Vente
A;Title: Whole-genome random sequencing and assembly of Haemophilus influenzae Rd.
A;Reference number: A64000; MUID:95350630
A;Accession: C64057

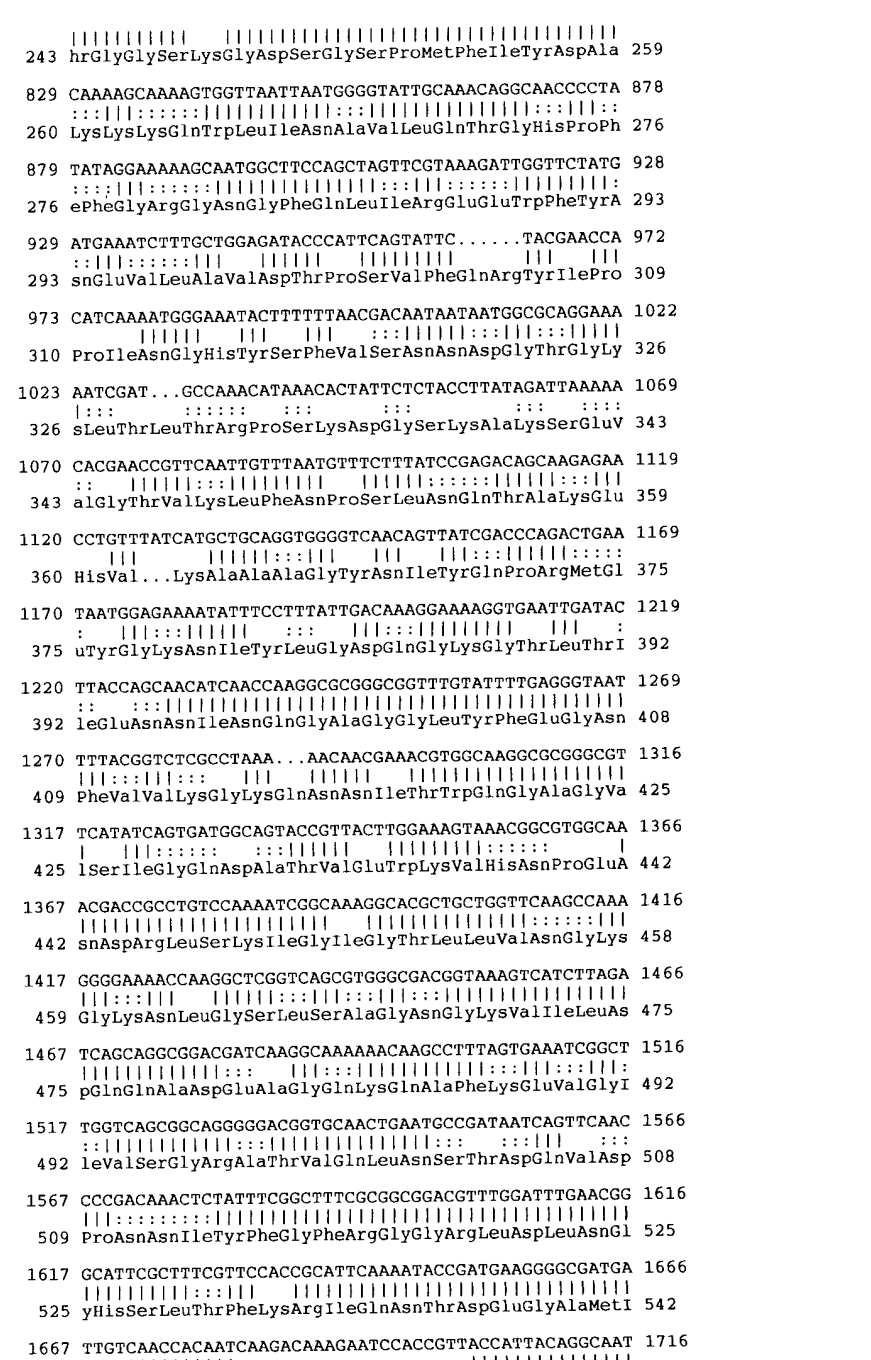
A;Status: nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-709 <TIGR>
A;Cross-references: GB:U32710; GB:L42023; NID:g1573200; PID:g1573211; TIGR:HI0248

alignment_scores:
Quality: 2494.00 Length: 701
Ratio: 4.075 Gaps: 5
Percent Similarity: 87.304 Percent Identity: 67.618

alignment_block:
US-09-303-518D-653 x C64057 ..

Align seg 1/1 to: C64057 from: 1 to: 709





```
  542 leValAsnHisAsnThrThrGlnValAlaAsnIleThrIleThrGlyAsn 558

 1717 AAAGATATTACTACAACCGGCAATAAC.........AACAACTTGGATAG 1757
      ::::::||||||:::   :::|||:::         |||:::||||||
  559 GluSerIleThrAlaProSerAsnLysLysAsnIleAsnLysLeuAspTy 575

 1758 CAAAAAAGAAATTGCCTACAACGGTTGGTTTGGCGAGAAAGATGCAACCA 1807
       :::||||||||||||||||||||||||||||||||   |||   :::|
  575 rSerLysGluIleAlaTyrAsnGlyTrpPheGlyGluThrAspLysAsnL 592

 1808 AAACGAACGGGCGGCTCAATCTGAATTACCAACCGGAAGAAGCGGATCGC 1857
      ||   ||||||||||||||||||   |||:::|||         ||||||
  592 ysHisAsnGlyArgLeuAsnLeuIleTyrLysProThrThrGluAspArg 608

 1858 ACTTTACTGCTTTCCGGCGGAACAAATTTAAACGGCAATATCACGCAAAC 1907
      ||||||||||||||||||||||||||||||:::|||:::||||||||||
  609 ThrLeuLeuLeuSerGlyGlyThrAsnLeuLysGlyAspIleThrGlnTh 625

 1908 AAACGGCAAACTGTTTTTCAGCGGCAGACCGACACCGCACGCCTACAATC 1957
      |:::||||||||||||||||||||||||||||||||||||||||||||||
  625 rLysGlyLysLeuPhePheSerGlyArgProThrProHisAlaTyrAsnH 642

 1958 ATTTAGGAAGCGGGTGGTCAAAAATGGAAGGTATCCCACAAGGAGAAATC 2007
      |||||   :::   ||||||:::|||||||||||||||||||||||||||
  642 isLeuAspLysArgTrpSerGluMetGluGlyIleProGlnGlyGluIle 658

 2008 GTGTGGGACAACGATTGGATCGACCGCACATTTAAAGCGGAAAACTTCCA 2057
      |||||||||   |||||||||:::|||||||||||||||||||||||::
  659 ValTrpAspTyrAspTrpIleAsnArgThrPheLysAlaGluAsnPheGl 675

 2058 TATTCAGGGCGGACAAGCGGTGGTTTCCCGCAATGTTGCCAAAGTGGAAG 2107
      :|||:::||||||:::|||||||||||||||||||||:::::::::||||
  675 nIleLysGlyGlySerAlaValValSerArgAsnValSerSerIleGluG 692

 2108 GCGATTGGCATTTAAGCAATCACGCCCAAGCAGTTTTCGGTGTCGCACCG 2157
      ||:::|||   :::||||||:::|||:::|||:::||||||||||:::|||
  692 lyAsnTrpThrValSerAsnAsnAlaAsnAlaThrPheGlyValValPro 708

 2158 CAT 2160
      :::
  709 Asn 709
```

seq_name: pir2:A37023

seq_documentation_block:
IgA-specific metalloendopeptidase (EC 3.4.24.13) type 1 precursor - Haemophilus influenz
N;Alternate names: immunoglobulin A1 proteinase type 1
C;Species: Haemophilus influenzae
C;Date: 31-Jan-1992 #sequence_revision 12-Jun-1992 #text_change 08-Dec-2000
C;Accession: A37023
R;Poulsen, K.; Brandt, J.; Hjorth, J.P.; Thogersen, H.C.; Kilian, M.
Infect. Immun. 57, 3097-3105, 1989
A;Title: Cloning and sequencing of the immunoglobulin A1 protease gene (iga) of Haemophi
A;Reference number: A37023; MUID:89379374
A;Accession: A37023
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1541 <POU>
A;Cross-references: GB:X64357; NID:g43560; PIDN:CAA45708.1; PID:g43561
A;Experimental source: serotype b
C;Superfamily: IgA-specific metalloendopeptidase
C;Keywords: hydrolase; metalloproteinase

alignment_scores:
Quality: 1614.00 Length: 1716
Ratio: 1.732 Gaps: 48
Percent Similarity: 54.312 Percent Identity: 26.865

alignment_block:
US-09-303-518D-653 x A37023 ..

Align seg 1/1 to: A37023 from: 1 to: 1541

```
   64 CGCTTCTCGCCCGCTTACTTAGCCATATGCCTGTCGTTCGGCATTCTGCC 113
      :::|||:::      :::::::|||:::   ::::::::::::::   ||
    5 LysPheLysLeuAsnPheIleAlaLeuThrValAlaTyrAlaLeuThrPr 21

  114 CCAAGCCCGGGCGGGACACACTTATTTCGGCATCAACTACCAATACTATC 163
      |   :::::::|||:::   :::         ::::::||||||   :::|
   21 oTyrThrGluAlaAlaLeuValArgAspAspValAspTyrGlnIlePheA 38

  164 GCGACTTTGCCGAAAATAAAGGCAAGTTTGCAGTCGGGGCGAAAGATATT 213
      ||||||||||||||||||||||||||||:::|||||||||   ::::::
   38 rgAspPheAlaGluAsnLysGlyLysPheSerValGlyAlaThrAsnVal 54

  214 GAGGTTTACAACAAAAAAGGGGAGTTGGTCGGCAAATCGATGACGAAAGC 263
         |||   :::|||::::::::::   :::|||   ::::::   :::::
   55 LeuValLysAspLysAsnAsnLysAspLeuGlyThrAlaLeuProAsnGl 71

  264 C...CCGATGATTGATTTTTCTGTGGTATCG...CGTAACGGCGTGGCGG 307
      :   ||||||||||||||||||||||:::       :::   ::::|||:
   71 yIleProMetIleAspPheSerValValAspValAspLysArgIleAlaT 88

  308 CATTGGCGGGCGATCAATATATTGTGAGCGTGGCACAT...AACGGCGGC 354
      ::|||   :::   ||||||:::|||:::|||   |||   ::::::|||
   88 hrLeuIleAsnProGlnTyrValValGlyValLysHisValSerAsnGly 104

  355 TATAACAATGTTGATTTTGGT..............GCGGAGGGAAGCAA 389
         :::::::::   ||||||              ::::::::::||
  105 ValSerGluLeuHisPheGlyAsnLeuAsnGlyAsnMetAsnAsnGlyAs 121

  390 TCCCGATCAGCACCGC..................TTTTCTTACCAAATTG 421
      |        ||||||                        |||      |
  121 nAlaLysAlaHisArgAspValSerSerGluGluAsnArgTyrPheSerV 138

  422 TGAAAAGAAATAATTATAAAGCAGGGACTAACGGC.............. 456
      ||:::::::|||:::|||   :::      ||||||
  138 alGluLysAsnGluTyrProThrLysLeuAsnGlyLysThrValThrThr 154

  457 ............CATCCTTATGGCGGCGATTATCATATGCCGCGTTTGCA 494
                  :::            ||||||:::||||||||||||
  155 GluAspGlnThrGlnLysArgArgGluAspTyrTyrMetProArgLeuAs 171

  495 CAAATTTGTCACAGATGCAGAACCTGTTGAG...ATGACCAGTTATATGG 541
       ||||||||||||:::::::   |||:::|||      |||:::      |
  171 pLysPheValThrGluValAlaProIleGluAlaSerThrAlaSerSerA 188

  542 ATGGGTGGAAATACGCTGATTTAAATAAATACCCTGATCGTGTTCGAATC 591
      ||:::      |||   |||   ||||||||||||      |||||||:::
  188 spAlaGlyThrTyrAsnAspGlnAsnLysTyrProAlaPheValArgLeu 204

  592 GGAGCAGGCAGACAATATTGGCGGTCTGATGAAGACGAA........... 630
      |||:::|||   |||:::      :::      |||:::
  205 GlySerGlySerGlnPheIleTyrLysLysGlyAspAsnTyrSerLeuIl 221

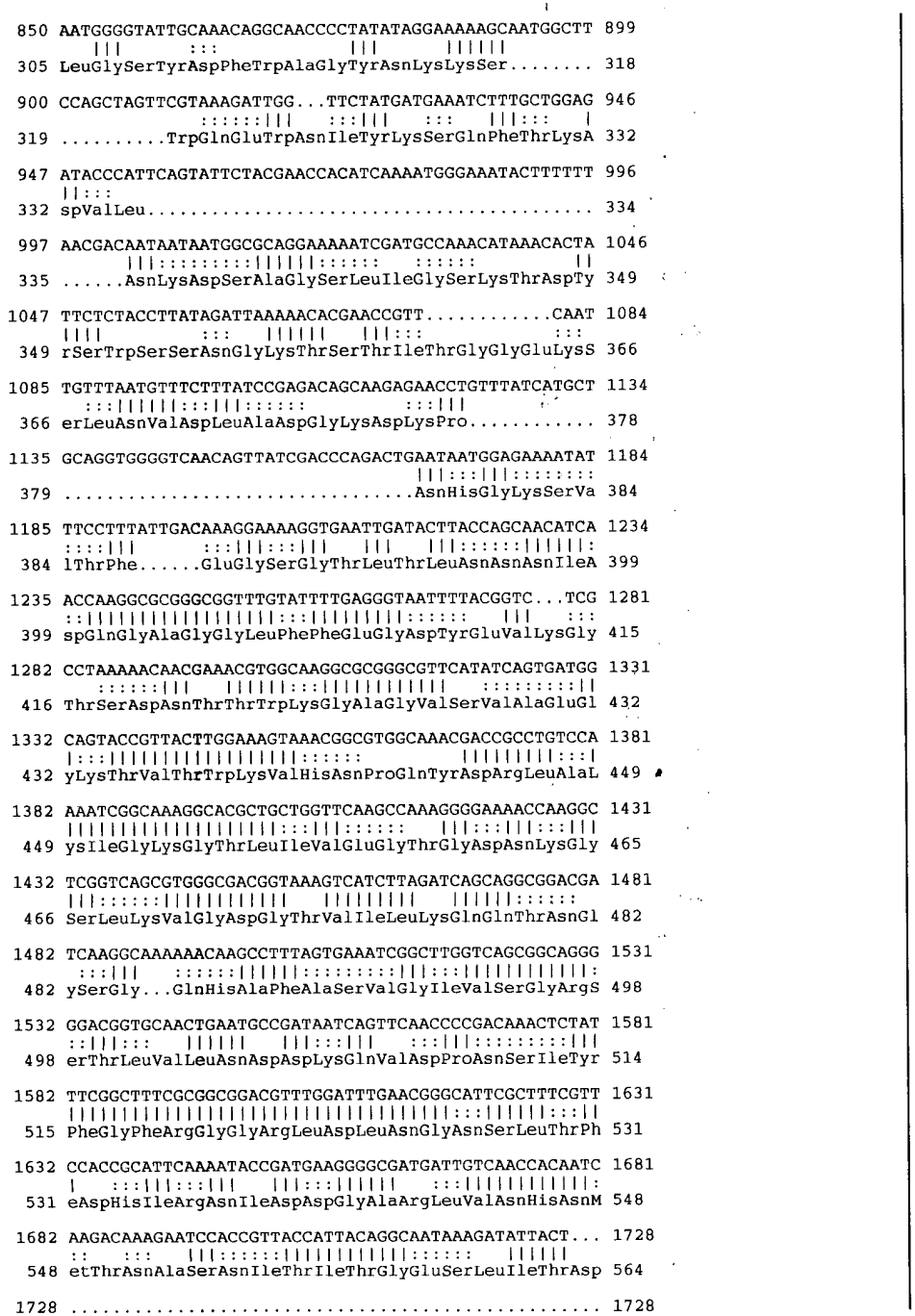
  631 .CCCAATAACCGCGAAAGTTCATATCAT............ATTGCAAGCG 667
       ||||||:::|||   :::   :::            :::::::::|
  221 eLeuAsnAsnHisGluValGlyGlyAsnAsnLeuLysLeuValGlyAspA 238

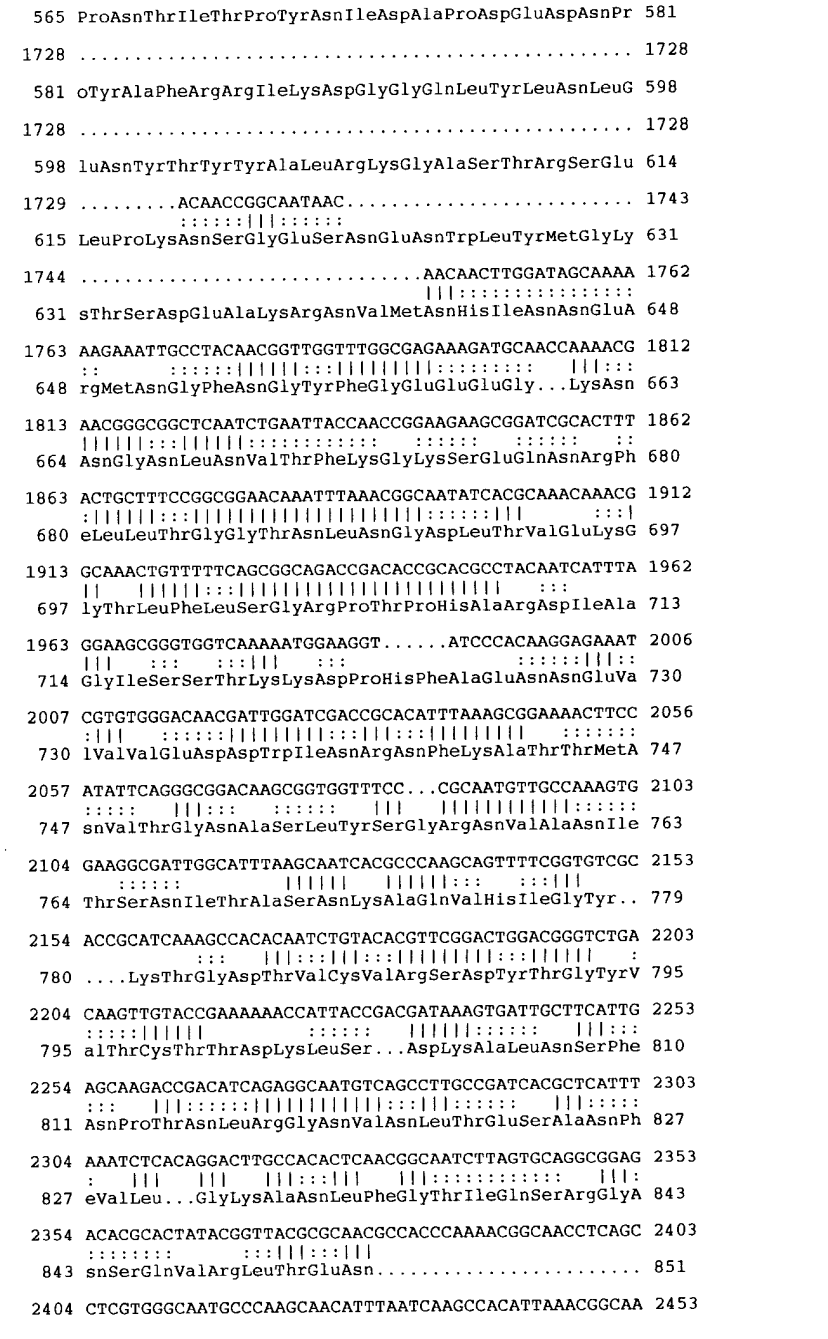
  668 CATATTCTTGGCTCGTCGGTGGCAATACCTTTGCACAAAATGGATCAGGT 717
      |||||::::::   ::::::|||:::   :::      |||   ::::::
  238 laTyrThrTyrGlyIleAlaGlyThrProTyrLysValAsnHisGluAsn 254

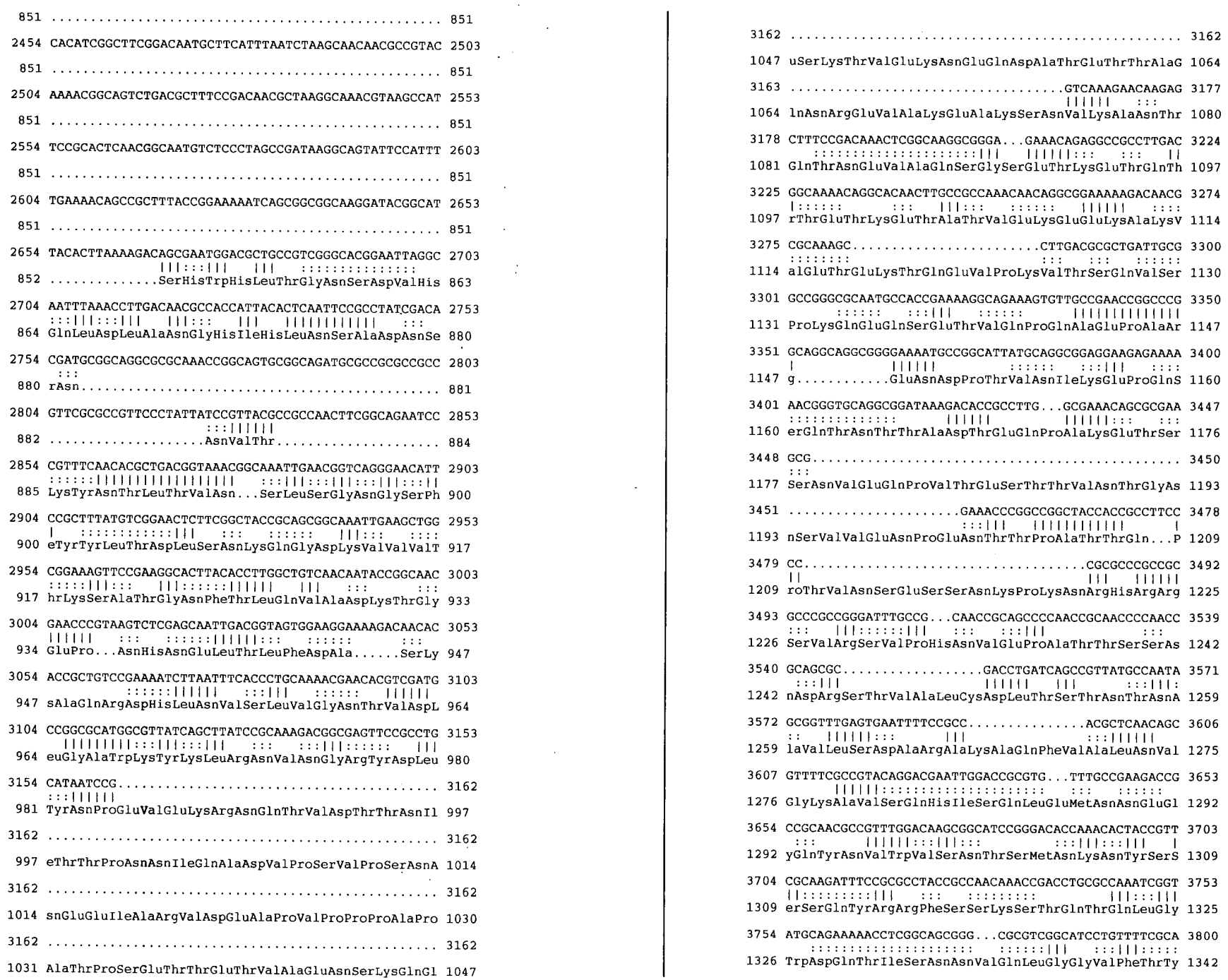
  718 GGTGGCACAGTCAACTTAGGTAGCGAAAAAATTAAACATAGC...CCATA 764
      :::|||   :::::::::|||:::::|||   :::||||||   |||
  255 AsnGlyLeuIleGlyPheGlyAsnSerLysGluGluHisSerAspProLy 271

  765 TGGTTTTTTACCAACAGGA..............GGCTCATTTGGCGACA 799
       |||:::|||                      :::   :::|||||||
  271 sGlyIleLeuSerGlnAspProLeuThrAsnTyrAlaValLeuGlyAspS 288

  800 GTGGCTCACCAATGTTTATCTATGATGCCCAAAAGCAAAAGTGGTTAATT 849
      ||||||||||:::|||:::||||||   :::|||   ||||||||:::
  288 erGlySerProLeuPheValTyrAspArgGluLysGlyLysTrpLeuPhe 304
```


850 AATGGGGTATTGCAAACAGGCAACCCCTATATAGGAAAAAGCAATGGCTT 899
305 LeuGlySerTyrAspPheTrpAlaGlyTyrAsnLysLysSer..... 318
900 CCAGCTAGTTCGTAAAGATTGG...TTCTATGATGAAATCTTTGCTGGAG 946
319 .....TrpGlnGluTrpAsnIleTyrLysSerGlnPheThrLysA 332
947 ATACCCATTCAGTATTCTACGAACCACATCAAAATGGGAAATACTTTTTT 996
332 spValLeu..... 334
997 AACGACAATAATAATGGCGCAGGAAAAATCGATGCCAAACATAAACACTA 1046
335 .....AsnLysAspSerAlaGlySerLeuIleGlySerLysThrAspTy 349
1047 TTCTCTACCTTATAGATTAAAAACACGAACCGTT.....CAAT 1084
349 rSerTrpSerSerAsnGlyLysThrSerThrIleThrGlyGlyGluLysS 366
1085 TGTTTAATGTTTCTTTATCCGAGACAGCAAGAGAACCTGTTTATCATGCT 1134
366 erLeuAsnValAspLeuAlaAspGlyLysAspLysPro..... 378
1135 GCAGGTGGGGTCAACAGTTATCGACCCAGACTGAATAATGGAGAAAATAT 1184
379 .....AsnHisGlyLysSerVa 384
1185 TTCCTTTATTGACAAAGGAAAAGGTGAATTGATACTTACCAGCAACATCA 1234
384 lThrPhe.....GluGlySerGlyThrLeuThrLeuAsnAsnAsnIleA 399
1235 ACCAAGGCGCGGGCGGTTTGTATTTTGAGGGTAATTTTACGGTC...TCG 1281
399 spGlnGlyAlaGlyGlyLeuPhePheGluGlyAspTyrGluValLysGly 415
1282 CCTAAAAACAACGAAACGTGGCAAGGCGCGGGCGTTCATATCAGTGATGG 1331
416 ThrSerAspAsnThrThrTrpLysGlyAlaGlyValSerValAlaGluGl 432
1332 CAGTACCGTTACTTGGAAAGTAAACGGCGTGGCAAACGACCGCCTGTCCA 1381
432 yLysThrValThrTrpLysValHisAsnProGlnTyrAspArgLeuAlaL 449
1382 AAATCGGCAAAGGCACGCTGCTGGTTCAAGCCAAAGGGGAAAACCAAGGC 1431
449 ysIleGlyLysGlyThrLeuIleValGluGlyThrGlyAspAsnLysGly 465
1432 TCGGTCAGCGTGGGCGACGGTAAAGTCATCTTAGATCAGCAGGCGGACGA 1481
466 SerLeuLysValGlyAspGlyThrValIleLeuLysGlnGlnThrAsnGl 482
1482 TCAAGGCAAAAAACAAGCCTTTAGTGAAATCGGCTTGGTCAGCGGCAGGG 1531
482 ySerGly...GlnHisAlaPheAlaSerValGlyIleValSerGlyArgS 498
1532 GGACGGTGCAACTGAATGCCGATAATCAGTTCAACCCCGACAAACTCTAT 1581
498 erThrLeuValLeuAsnAspAspLysGlnValAspProAsnSerIleTyr 514
1582 TTCGGCTTTCGCGGCGGACGTTTGGATTTGAACGGGCATTCGCTTTCGTT 1631
515 PheGlyPheArgGlyGlyArgLeuAspLeuAsnGlyAsnSerLeuThrPh 531
1632 CCACCGCATTCAAAATACCGATGAAGGGGCGATGATTGTCAACCACAATC 1681
531 eAspHisIleArgAsnIleAspAspGlyAlaArgLeuValAsnHisAsnM 548
1682 AAGACAAAGAATCCACCGTTACCATTACAGGCAATAAAGATATTACT... 1728
548 etThrAsnAlaSerAsnIleThrIleThrGlyGluSerLeuIleThrAsp 564
1728 1728


565 ProAsnThrIleThrProTyrAsnIleAspAlaProAspGluAspAsnPr 581
1728 1728
581 oTyrAlaPheArgArgIleLysAspGlyGlyGlnLeuTyrLeuAsnLeuG 598
1728 1728
598 luAsnTyrThrTyrTyrAlaLeuArgLysGlyAlaSerThrArgSerGlu 614
1729 .....ACAACCGGCAATAAC..... 1743
615 LeuProLysAsnSerGlyGluSerAsnGluAsnTrpLeuTyrMetGlyLy 631
1744 .....AACAACTTGGATAGCAAAA 1762
631 sThrSerAspGluAlaLysArgAsnValMetAsnHisIleAsnAsnGluA 648
1763 AAGAAATTGCCTACAACGGTTGGTTTGGCGAGAAAGATGCAACCAAAACG 1812
648 rgMetAsnGlyPheAsnGlyTyrPheGlyGluGluGluGly...LysAsn 663
1813 AACGGGCGGCTCAATCTGAATTACCAACCGGAAGAAGCGGATCGCACTTT 1862
664 AsnGlyAsnLeuAsnValThrPheLysGlyLysSerGluGlnAsnArgPh 680
1863 ACTGCTTTCCGGCGGAACAAATTTAAACGGCAATATCACGCAAACAAACG 1912
680 eLeuLeuThrGlyGlyThrAsnLeuAsnGlyAspLeuThrValGluLysG 697
1913 GCAAACTGTTTTTCAGCGGCAGACCGACACCGCACGCCTACAATCATTTA 1962
697 lyThrLeuPheLeuSerGlyArgProThrProHisAlaArgAspIleAla 713
1963 GGAAGCGGGTGGTCAAAAATGGAAGGT.....ATCCCACAAGGAGAAAT 2006
714 GlyIleSerSerThrLysLysAspProHisPheAlaGluAsnAsnGluVa 730
2007 CGTGTGGGACAACGATTGGATCGACCGCACATTTAAAGCGGAAAACTTCC 2056
730 lValValGluAspAspTrpIleAsnArgAsnPheLysAlaThrThrMetA 747
2057 ATATTCAGGGCGGACAAGCGGTGGTTTCC...CGCAATGTTGCCAAAGTG 2103
747 snValThrGlyAsnAlaSerLeuTyrSerGlyArgAsnValAlaAsnIle 763
2104 GAAGGCGATTGGCATTTAAGCAATCACGCCCAAGCAGTTTTCGGTGTCGC 2153
764 ThrSerAsnIleThrAlaSerAsnLysAlaGlnValHisIleGlyTyr.. 779
2154 ACCGCATCAAAGCCACACAATCTGTACACGTTCGGACTGGACGGGTCTGA 2203
780 ....LysThrGlyAspThrValCysValArgSerAspTyrThrGlyTyrV 795
2204 CAAGTTGTACCGAAAAAACCATTACCGACGATAAAGTGATTGCTTCATTG 2253
795 alThrCysThrThrAspLysLeuSer...AspLysAlaLeuAsnSerPhe 810
2254 AGCAAGACCGACATCAGAGGCAATGTCAGCCTTGCCGATCACGCTCATTT 2303
811 AsnProThrAsnLeuArgGlyAsnValAsnLeuThrGluSerAlaAsnPh 827
2304 AAATCTCACAGGACTTGCCACACTCAACGGCAATCTTAGTGCAGGCGGAG 2353
827 eValLeu...GlyLysAlaAsnLeuPheGlyThrIleGlnSerArgGlyA 843
2354 ACACGCACTATACGGTTACGCGCAACGCCACCCAAAACGGCAACCTCAGC 2403
843 snSerGlnValArgLeuThrGluAsn..... 851
2404 CTCGTGGGCAATGCCCAAGCAACATTTAATCAAGCCACATTAAACGGCAA 2453



```
 3801 CAACCGGACCGGAAACACCTTCGACGACGGCATCGGCAACTCGGCACGGC 3850
      :   |||:::::|||:::|||||   :::   ::::::::::::
 1342 rValArgAsnSerAsnAsnPheAspLysAlaThrSerLysAsnThrLeuA 1359

 3851 TTGCCCACGGTGCCGTTTTCGGGCAATACGGCATCGGCAGGTTCGACATC 3900
             :::               |||:::   :::::::::   :::
 1359 laGlnValAsnPheTyrSerLysTyrTyrAlaAspAsnHisTrpTyrLeu 1375

 3901 GGCATCAGCGCGGGCGCGGGTTTTAGTAGCGGCAGCCTTTCAGACGGCAT 3950
      |||||||:::   |||   |||      ::::::::::|||:::   :::
 1376 GlyIleAspLeuGlyTyrGlyLysPheGlnSerLysLeuGlnThrAsnHi 1392

 3951 CAGAGGCAAAATCCGCCGCCGCGTGCTGCATTACGGCATTCAGGCAAGAT 4000
       :::::|||:::   |||::::::   ::::::|||:::   |||
 1392 sAsnAlaLysPheAlaArgHisThrAlaGlnPheGlyLeuThrAlaGlyL 1409

 4001 ACCGCGCAGGTTTCGGCGGATTCGGCATCGAACCGCACATCGGCGCAACG 4050
             ::::::|||:::|||||||||   |||   :::|||:::
 1409 ysAlaPheAsnLeuGlyAsnPheGlyIleThrProIleValGlyValArg 1425

 4051 CGCTATTTCGTCCAAAAAGCGGATTACCGATACGAAAACGTCAATATCGC 4100
            :::::::::::::||||||:::      :::::::::::|||
 1426 TyrSerTyrLeuSerAsnAlaAspPheAlaLeuAspGlnAlaArgIleLy 1442

 4101 CACCCCGGGCCTTGCATTCAACCGCTACCGCGCGGGCATTAAGGCAGATT 4150
       :::      ::::::   :::         |||   :::      :::|
 1442 sValAsnProIleSerValLysThrAlaPheAlaGlnValAspLeuSerT 1459

 4151 ATTCATTCAAACCGGCGCAACACATTTCCATCACGCCTTATTTGAGCCTG 4200
      ||::::::      ::::::   :::|||:::||||||   ||||||
 1459 yrThrTyrHisLeuGlyGlu...PheSerValThrProIleLeuSerAla 1474

 4201 TCCTATACCGATGCCGCTTCCGGCAAAGTCCGAACGCGCGTCAATACCGC 4250
        |||   ||||||   :::|||:::   :::   :::||||||
 1475 ArgTyr...AspAlaAsnGlnGlySerGlyLysIleAsnValAsnGlyTy 1490

 4251 CGTATTGGCGCAGGATTTCGGCAAAACCCGCAGTGCGGAATGGGGCGTAA 4300
          :::|||   :::      :::   ::::::   :::   |||::::
 1490 rAspPheAlaTyrAsnValGluAsnGlnGlnGlnTyrAsnAlaGlyLeuL 1507

 4301 ACGCCGAAATCAAAGGTTTCACGCTGTCCCTCCACGCTGCCGCCGCCAAG 4350
      ::   :::      :::      |||||||||   ::::::   :::|||
 1507 ysLeuLysTyrHisAsnValLysLeuSerLeuIleGlyGlyLeuThrLys 1523

 4351 GGGCCGCAATTGGAAGCGCAGCACAGCGCGGGCATCAAATTAGGCTAC 4398
      :::   |||   |||   |||   :::|||   :::||||||::::::
 1524 AlaLysGlnAlaGluLysGlnLysThrAlaGluLeuLysLeuSerPhe 1539

seq_name: pir2:S61314

seq_documentation_block:
IgA-specific metalloendopeptidase (EC 3.4.24.13) homolog - Neisseria meningitidis
C;Species: Neisseria meningitidis
A;Variety: HF13
C;Date: 20-Jul-1996 #sequence_revision 13-Mar-1997 #text_change 08-Dec-2000
C;Accession: S61314
R;Lomholt, H.; Poulsen, K.; Mogens, K.
Mol. Microbiol. 15, 495-506, 1995
A;Title: Comparative characterization of the iga gene encoding IgA1 protease in Neisseri
A;Reference number: S61314; MUID:95302961
A;Accession: S61314
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1561 <LOM>
A;Cross-references: EMBL:X82474; NID:g732873; PIDN:CAA57857.1; PID:g732874
C;Superfamily: IgA-specific metalloendopeptidase
C;Keywords: hydrolase; metalloproteinase


alignment_scores:
           Quality: 1605.00            Length:   1749
             Ratio:  1.750              Gaps:     49
 Percent Similarity:  52.430   Percent Identity: 27.616
```

```
alignment_block:
 US-09-303-518D-653 x S61314   ..

 Align seg 1/1  to: S61314  from: 1  to: 1561

      49 AAAACCGGCCGCATCCGCTTCTCGCCCGCTTACTTAGCCATATGCCTGTC 98
         ||||||   |||:::::::::::::         |||:::|||   |||::
       2 LysThrLysArgPheLysIleAsnAlaIleSerLeuSerIlePheLeuAl 18

      99 GTTCGGCATTCTGCCCCAAGCCCGGGCGGGACACACTTATTTCGGCATCA 148
         ::::::::::   |||   ::::::|||:::   :::         ::::
      18 aTyrAlaLeuThrProTyrSerGluAlaAlaLeuValArgAspAspValA 35

     149 ACTACCAATACTATCGCGACTTTGCCGAAAATAAAGGCAAGTTTGCAGTC 198
         ::||||||   :::||||||||||||||||||||||||||||||   |||
      35 spTyrGlnIlePheArgAspPheAlaGluAsnLysGlyLysPhePheVal 51

     199 GGGGCGAAAGATATTGAGGTTTACAACAAAAAAGGGGAGTTGGTCGGCAA 248
         ||||||   |||:::::::|||   ||||||:::|||:::   :::|||::
      52 GlyAlaThrAspLeuSerValLysAsnLysGlnGlyGlnAsnIleGlyAs 68

     249 ATCGATGACGAAAGCCCCGATGATTGATTTTTCTGTGGTATCGCGTAAC. 297
         ::::::::::::::::|||||||||||||||||||||:::::   |||
      68 nAlaLeuSerAsnValProMetIleAspPheSerValAlaAspValAsnL 85

     298 ..GGCGTGGCGGCATTGGCGGGCGATCAATATATTGTGAGCGTGGCACAT 345
            :::|||:::::::::      ||||||   |||||||||   |||
      85 ysArgIleAlaThrValValAspProGlnTyrAlaValSerValLysHis 101

     346 ..........................AACGGCGGCTATAACAATGTTGA 368
                                    |||   ||||||:::   ::
     102 AlaLysAlaGluValHisThrPheTyrTyrGlyGlnTyrAsnGlyHisAs 118

     369 TTTTGGTGCGGAGGGAAGCAATCCCGATCAGCACCGCTTTTCTTACCAAA 418
         :      |||:::   :::|||                   :::|||::::
     118 nAspValAlaAspLysGluAsn.................GluTyrArgV 129

     419 TTGTGAAAAGAAATAATTATAAAGCAGGGACTAACGGCCATCCTTATGGC 468
         ::|||::::::||||||||:::                    |||:::
     129 alValGluGlnAsnAsnTyrGlu.................ProHisLys 139

     469 GGC............................GATTATCATATGCCGCG 488
         :::                            ||||||:::|||   ||
     140 AlaTrpSerAlaSerAsnLeuGlyArgLeuGluAspTyrAsnMetAlaAr 156

     489 TTTGCACAAATTTGTCACAGATGCAGAACCTGTTGAGATGACCAGTTATA 538
         |:::::|||||||||||||:::::   |||:::      |||:::
     156 gPheAsnLysPheValThrGluValAlaProIleAlaProThrAspAlaG 173

     539 TGGATGGGTGG...AAATACGCTGATTTAAATAAATACCCTGATCGTGTT 585
              |||         |||   |||   |||::::::   :::   |||
     173 lyGlyGlyLeuAspThrTyrLysAspLysAsnArgPheSerSerPheVal 189

     586 CGAATCGGAGCAGGCAGACAA......TATTGGCGGTCTGATGAAGACGA 629
         |||:::|||||||||||||||      |||   ::::::      ::
     190 ArgValGlyAlaGlyArgGlnLeuValTyrGluLysGlyAlaTyrHisGl 206

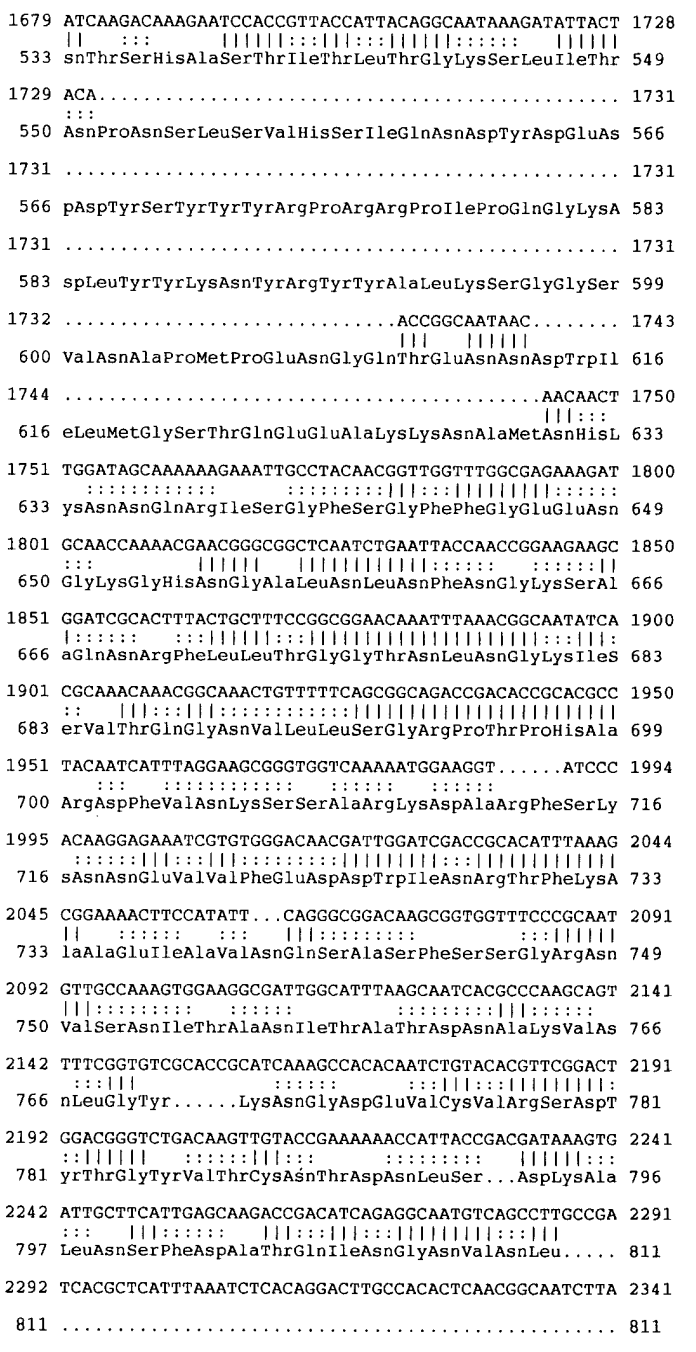
     630 ACCCAATAACCGCGAAAGTTCATAT.........CATATTGCAAGCGCAT 670
         :   :::|||   |||:::::::|||           ::::::::::||||
     206 nGluGlyAsn...GluLysGlyTyrAspLeuArgAspLeuSerGlnAlaT 222

     671 ATTCTTGGCTCGTCGGTGGC............................. 690
         ||   :::   ::::::|||
     222 yrArgTyrAlaIleAlaGlyThrProTyrLysAspIleAsnIleAspGln 238

     691 ......AATACCTTTGCACAAAATGGATCAGGTGGTGGCACAGTCAACTT 734
               ||||||   :::      |||   |||:::   :::::::::
     239 ThrMetAsnThrGluGlyLeuIleGlyPheGlyAsnHisAsnThrHisTy 255

     735 AGGTAGCGAAAAAATTAAACATAGC...CCATATGGTTTTTTACCAACAG 781
          :::::::|||:::::::|||::::::               |||   :::
     255 rSerAlaGluGluLeuLysGlnAlaLeuSerGlnAspAlaLeuThrAsnT 272
```


782 GAGGCTCATTTGGCGACAGTGGCTCACCAATGTTTATCTATGATGCCCAA 831
272 yrGlyValLeuGlyAspSerGlySerProLeuPheAlaPheAspLysGln 288
832 AAGCAAAAGTGGTTAATTAATGGGGTATTGCAAACAGGCAACCCCTATAT 881
289 LysAsnGlnTrpValPheLeuGlyThrTyrAspTyrTrpAlaGlyTyrGl 305
882 AGGAAAAAGCAATGGCTTCCAGCTAGTTCGTAAAGATTGGTTCTATGATG 931
305 yLysLysSerTrpGlnGluTrpAsnIleTyrLysLysGluPheAlaAspL 322
932 AAATCTTTGCTGGAGATACCCATTCAGTATTCTACGAACCACATCAAAAT 981
322 ysIle..... 323
982 GGGAAATACTTTTTTAACGACAATAATAATGGCGCAGGAAAAATCGATGC 1031
324 .....LysGlnArgAspAsnAlaGlyThrIleLysGl 334
1032 CAAACATAAACACTATTCTCTACCTTATAGATTAAAAACACGAACCGTTC 1081
334 yTyrGlyGluHisHis.....TrpLysThrThrGlyThr. 345
1082 AATTGTTTAATGTTTCTTTATCCGAGACAGCAAGAGAACCTGTTTATCAT 1131
346 .....AsnSerHisIleGlySerThr..... 352
1132 GCTGCAGGTGGGGTCAACAGTTATCGACCCAGACTGAATAATGGAGAAAA 1181
353 AlaValArgLeuAlaAsnAsnGluArgAsp...AlaAsnAsnGlyGlnAs 368
1182 TATTTCCTTTATTGACAAAGGAAAAGGTGAATTGATACTTACCAGCAACA 1231
368 nValThrPhe.....GluAsnAsnGlyThrLeuValLeuAspGlnAsnI 383
1232 TCAACCAAGGCGCGGGCGGTTTGTATTTTGAGGGTAATTTTACGGTCTCG 1281
383 leAsnGlnGlyAlaGlyGlyLeuPhePheLysGlyAspTyrThrValLys 399
1282 CCTAAAAACAACGAA...ACGTGGCAAGGCGCGGGCGTTCATATCAGTGA 1328
400 GlyAlaAsnAsnAspIleThrTrpLeuGlyAlaGlyIleAspValAlaAs 416
1329 TGGCAGTACCGTTACTTGGAAAGTAAACGGCGTGGCAAACGACCCGCCTGT 1378
416 pGlyLysLysValValTrpGlnValLysAsnProAsnGlyAspArgLeuA 433
1379 CCAAAATCGGCAAAGGCACGCTGCTGGTTCAAGCCAAAGGGGAAAACCAA 1428
433 laLysIleGlyLysGlyThrLeuGluIleAsnGlyThrGlyValAsnGln 449
1429 GGCTCGGTCAGCGTGGGCGACGGTAAAGTCATCTTAGATCAGCAGGCGGA 1478
450 GlyGluLeuLysValGlyAspGlyThrValIleLeuAsnGlnLysAlaAs 466
1479 CGATCAAGGCAAAAAACAAGCCTTTAGTGAAATCGGCTTGGTCAGCGGCA 1528
466 pSerAsnGlnLysValSerAlaPheSerGlnValGlyIleValSerGlyA 483
1529 GGGGGACGGTGCAACTGAATGCCGATAATCAGTTCAACCCCGACAAACTC 1578
483 rgGlyThrLeuValLeuAsnSerSerAsnGlnIleAsnProAspAsnLeu 499
1579 TATTTCGGCTTTCGCGGCGGACGTTTGGATTTGAACGGGCATTCGCTTTC 1628
500 TyrPheGlyPheArgGlyGlyArgLeuAspAlaAsnGlyAsnAspLeuTh 516
1629 GTTCCACCGCATTCAAAATACCGATGAAGGGGCGATGATTGTCAACCACA 1678
516 rPheGluHisIleArgAsnValAspGluGlyAlaArgIleValAsnHisA 533


1679 ATCAAGACAAAGAATCCACCGTTACCATTACAGGCAATAAAGATATTACT 1728
533 snThrSerHisAlaSerThrIleThrLeuThrGlyLysSerLeuIleThr 549
1729 ACA..... 1731
550 AsnProAsnSerLeuSerValHisSerIleGlnAsnAspTyrAspGluAs 566
1731 1731
566 pAspTyrSerTyrTyrTyrArgProArgArgProIleProGlnGlyLysA 583
1731 1731
583 spLeuTyrTyrLysAsnTyrArgTyrTyrAlaLeuLysSerGlyGlySer 599
1732 .....ACCGGCAATAAC..... 1743
600 ValAsnAlaProMetProGluAsnGlyGlnThrGluAsnAsnAspTrpIl 616
1744 .....AACAACT 1750
616 eLeuMetGlySerThrGlnGluGluAlaLysLysAsnAlaMetAsnHisL 633
1751 TGGATAGCAAAAAAGAAATTGCCTACAACGGTTGGTTTGGCGAGAAAGAT 1800
633 ysAsnAsnGlnArgIleSerGlyPheSerGlyPhePheGlyGluGluAsn 649
1801 GCAACCAAAACGAACGGGCGGCTCAATCTGAATTACCAACCGGAAGAAGC 1850
650 GlyLysGlyHisAsnGlyAlaLeuAsnLeuAsnPheAsnGlyLysSerAl 666
1851 GGATCGCACTTTACTGCTTTCCGGCGGAACAAATTTAAACGGCAATATCA 1900
666 aGlnAsnArgPheLeuLeuThrGlyGlyThrAsnLeuAsnGlyLysIleS 683
1901 CGCAAACAAACGGCAAACTGTTTTTCAGCGGCAGACCGACACCGCACGCC 1950
683 erValThrGlnGlyAsnValLeuLeuSerGlyArgProThrProHisAla 699
1951 TACAATCATTTAGGAAGCGGGTGGTCAAAAATGGAAGGT.....ATCCC 1994
700 ArgAspPheValAsnLysSerSerAlaArgLysAspAlaArgPheSerLy 716
1995 ACAAGGAGAAATCGTGTGGGACAACGATTGGATCGACCGCACATTTAAAG 2044
716 sAsnAsnGluValValPheGluAspAspTrpIleAsnArgThrPheLysA 733
2045 CGGAAAACTTCCATATT...CAGGGCGGACAAGCGGTGGTTTCCCGCAAT 2091
733 laAlaGluIleAlaValAsnGlnSerAlaSerPheSerSerGlyArgAsn 749
2092 GTTGCCAAAGTGGAAGGCGATTGGCATTTAAGCAATCACGCCCAAGCAGT 2141
750 ValSerAsnIleThrAlaAsnIleThrAlaThrAspAsnAlaLysValAs 766
2142 TTTCGGTGTCGCACCGCATCAAAGCCACACAATCTGTACACGTTCGGACT 2191
766 nLeuGlyTyr.....LysAsnGlyAspGluValCysValArgSerAspT 781
2192 GGACGGGTCTGACAAGTTGTACCGAAAAAACCATTACCGACGATAAAGTG 2241
781 yrThrGlyTyrValThrCysAsnThrAspAsnLeuSer...AspLysAla 796
2242 ATTGCTTCATTGAGCAAGACCGACATCAGAGGCAATGTCAGCCTTGCCGA 2291
797 LeuAsnSerPheAspAlaThrGlnIleAsnGlyAsnValAsnLeu..... 811
2292 TCACGCTCATTTAAATCTCACAGGACTTGCCACACTCAACGGCAATCTTA 2341
811 811
2342 GTGCAGGCGGAGACACGCACTATACGGTTACGCGCAACGCCACCCAAAAC 2391

```
 812 ..........................................Asn 812

2392 GGCAACCTCAGCCTCGTGGGCAATGCCCAAGCAACATTTAATCAAGCCAC 2441
 813 GlnAsnAlaAlaLeuValLeuGlyLysAlaAlaLeuTrpGlyGlnIleGl 829

2442 ATTAAACGGCAACACATCGGCTTCGGACAATGCTTCATTTAATCTAAGCA 2491
 829 nGlyGlnGlyAsnSerSerValSerLeuAsn..................... 839

2492 ACAACGCCGTACAAAACGGCAGTCTGACGCTTTCCGACAACGCTAAGGCA 2541
 839 ..................................................... 839

2542 AACGTAAGCCATTCCGCACTCAACGGCAATGTCTCCCTAGCCGATAAGGC 2591
 839 ..................................................... 839

2592 AGTATTCCATTTTGAAAACAGCCGCTTTACCGGAAAAATCAGCGGCGGCA 2641
 839 ..................................................... 839

2642 AGGATACGGCATTACACTTAAAAGACAGCGAATGGACGCTGCCGTCGGGC 2691
 840 ...................GlnHisSerLysTrpHisLeuThrSerAsp 849

2692 ACGGAATTAGGCAATTTAAACCTTGACAACGCCACCATTACACTCAATTC 2741
 850 SerGlnValHisAsnLeuSerLeuAlaAspSerHisIleHisLeuAsnAs 866

2742 CGCCTATCGACACGATGCGGCAGGCGCGCAAACCGGCAGTGCGGCAGATG 2791
 866 nAla......SerAspAlaGlnSerAla..................... 873

2792 CGCCGCGCCGCCGTTCGCGCCGTTCCCTATTATCCGTTACGCCGCCAACT 2841
 873 ..................................................... 873

2842 TCGGCAGAATCCCGTTTCAACACGCTGACGGTAAACGGCAAATTGAACGG 2891
 874 .........AsnLysTyrHisThrLeuLysIleAsn...HisLeuSerGl 886

2892 TCAGGGAACATTCCGCTTTATGTCGGAACTCTTCGGCTACCGCAGCGGCA 2941
 886 yAsnGlyHisPheHisTyrLeuThrHisLeuAlaLysAsnLeuGlyAspL 903

2942 AATTGAAGCTGGCGGAAAGTTCCGAAGGCACTTACACCTTGGCTGTCAAC 2991
 903 ysValLeuValLysGluSerAlaSerGlyHisTyrGlnLeuHisValGln 919

2992 AATACCGGCAACGAACCCGTAAGTCTCGAGCAATTGACGGTAGTGGAAGG 3041
 920 AspLysThrGlyGluPro...AsnGlnGluGlyLeuAsnLeuPheAspAl 935

3042 AAAAGACAACACACCGCTGTCCGAAAATCTTAATTTCACCCTGCAAAACG 3091
 935 aSerSerValGlnAspArgSer...ArgLeuSerValSerLeuAlaAsnA 951

3092 AACACGTCGATGCCGGCGCATGGCGTTATCAGCTTATCCGCAAAGACGGC 3141
 951 snHisValAspLeuGlyAlaLeuArgTyrThrIleLysThrGluAsnGly 967

3142 GAGTTCCGCCTGCATAATCCG..................................... 3162
 968 IleThrArgLeuTyrAsnProTyrAlaGluAsnArgArgArgValLysPr 984

3162 ..................................................... 3162
 984 oAlaProSerProAlaThrAsnThrAlaSerGlnAlaGlnThrAspSerA 1001

3162 ..................................................... 3162
```

```
1001 laGlnIleAlaLysProGlnAsnIleValValAlaProProSerProGln 1017

3163 ..................GTCAAAGAACAAGAGCTTTCCGACAAACTCGG 3194
1018 AlaAsnGlnAlaGluGluAlaLysArgGlnGlnAlaLysAlaGluGlnVa 1034

3195 CAAGGCGGGAGAAACAGAGGCCGCCTTGACGGCA..................... 3228
1034 lLysArgGlnGlnAlaGluAlaGluLysValAlaHisGlnLysAlaGluG 1051

3228 ..................................................... 3228
1051 luAlaLysArgGlnGlnAspAlaLeuAlaArgGlnGlnAlaGluGlnGlu 1067

3229 .....................AAACAGGCACAACTTGCCGCCAAACAACA 3257
1068 ArgGlnArgLeuGluAlaGluArgGlnAlaAlaGluIleAlaLysGlnLy 1084

3258 GGCGGAAAAAGACAACGCGCAAAGCCTTGACGCGCTGATTGCGGCCGGGC 3307
1084 sAlaGluAlaGluGluAlaLysArgArgAlaAlaGluIleAlaGluGlnL 1101

3308 GCAATGCCACCGAAAAGGCAGAA...AGTGTTGCCGAACCGGCCCGGCAG 3354
1101 ysAlaAlaAlaGluGluAlaLysArgGlnAlaAlaGluLeuAlaArgGln 1117

3355 ..............GCAGGCGGGGAAAATGCCGGCATTATGCAGGCGGA 3389
1118 GlnGluGluAlaArgLysAlaAlaGluLeuAlaAlaLysGlnLysAlaGl 1134

3390 GGAAGAGAAAAAA.....................CGGGTGCAGGCGGATA 3418
1134 uThrGluArgLysAlaAlaGluIleAlaGluGlnLysAlaGluAlaGluA 1151

3419 AAGACACCGCC...TTGGCGAAACAGCGC...GAAGCGGAAACCCGGCCG 3462
1151 rgGluAlaAlaGluLeuAlaLysGlnLysAlaGluGluGluGlyArgGln 1167

3463 GCTACCACCGCCTTCCCCCGCGCCCGCCGCGCCCGCCGGGATTTGCCG.. 3510
1168 AlaAlaGlnSerGlnProLys...ArgArgAsnArgArgAlaIleProPr 1183

3510 ..................................................... 3510
1183 oGluLeuSerSerAspAlaThrThrArgAlaLeuProArgIleAlaArgA 1200

3510 ..................................................... 3510
1200 snSerAsnProAspAlaSerAspTyrGluGluIleProLeuAspAlaLeu 1216

3510 ..................................................... 3510
1217 GluAspGluAspValSerGluSerValAspThrSerAspLysGlnProGl 1233

3510 ..................................................... 3510
1233 nAspAsnThrGluLeuHisGluLysValGluThrValSerLeuGlnProA 1250

3511 ........CAACCG..................CAGCCCCAACCGCAACCC 3534
1250 rgAlaAlaGlnProArgAlaGlnAlaAlaAlaGlnProGlnAlaGlnAla 1266

3535 CAACCGCAGCGCGACCTGATCAGCCGTTATGCCAATAGCGGTTTGAGTGA 3584
1267 ValAlaGlnAlaAspAlaValSerThrAsnThrAsnSerAlaLeuSerAs 1283

3585 ATTTTCCGCCACGCTCAACAGCGTTTTCGCC...GTACAGGACGAATTGG 3631
1283 pAlaMetAlaSerThrGlnSerIleLeuLeuAspThrGlyAlaSerLeuT 1300

3632 ACCGCGTGTTTGCCGAAGACCGCCGC..............AACGCCGTT 3666
1300 hrArgHisIleAlaGlnLysSerArgAlaAspAlaGluLysAsnSerVal 1316
```

```
 3667 TGGACAAGCGGCATCCGGGACACCAAACACTACCGTTCGCAAGATTTCCG 3716
      |||   |||:::|||          :::   |||   |||   ::::::||
 1317 TrpMetSerAsnIleGlyTyrGlyArgAspTyrAlaSerAlaGlnTyrAr 1333

 3717 CGCCTACCGCCAACAAACCGACCTGCGCCAAATCGGTATGCAGAAAAACC 3766
      |   :::   ::::::          ||||||||||:::::::::::::|
 1333 gArgPheSerSerLysArgThrGlnThrGlnIleGlyIleAspArgSerL 1350

 3767 TCGGCAGCGGG...CGCGTCGGCATCCTGTTTTCGCACAACCGGACCGGA 3813
      ||:::::::::   :::::::|||   :::::::::::::   :::
 1350 euSerGluAsnMetGlnIleGlyGlyValLeuThrTyrSerAspSerGln 1366

 3814 AACACCTTCGACGACGGCATCGGCAACTCGGCACGGCTTGCCCACGGTGC 3863
      :::||||||||||:::::   |||:::::::::   ::::::::::::
 1367 HisThrPheAspGlnAlaSerGlyLysAsnThr...PheValGlnAlaAs 1382

 3864 CGTTTTCGGGCAATACGGCATCGGC...AGGTTCGACATCGGCATCAGCG 3910
      :::::::|||:::|||   ::::::      :::   ::::::   :::
 1382 nLeuTyrGlyLysTyrTyrLeuAsnAspAlaTrpTyrValAlaGlyAspI 1399

 3911 CGGGCGCGGGTTTTAGTAGCGGCAGCCTTTCAGACGGCATCAGAGGCAAA 3960
        |||||||||         :::   |||:::         :::::::::
 1399 leGlyAlaGlySerLeuArgSerArgLeuGlnThrGlnGlnLysAlaAsn 1415

 3961 ATCCGCCGCCGCGTGCTGCATTACGGCATTCAGGCAAGATACCGCGCAGG 4010
      :::::::|||      ::::::   |||:::
 1416 PheAsnArgThrSerIleGlnThrGlyLeuThrLeuGlyAsnThrLeuLy 1432

 4011 TTTCGGCGGATTCGGCATCGAACCGCACATCGGCGCAACGCGCTATTTCG 4060
       ::::::   |||   |||   |||       |||          :
 1432 sIleAsnGlnPheGluIleValProSerAlaGlyIleArgTyrSerArgL 1449

 4061 TCCAAAAAGCGGATTACCGATACGAAAACGTCAATATCGCCACCCCGGGC 4110
      ::::::::|||||||||:::      |||   ::::::   :::   :::
 1449 euSerSerAlaAspTyrLysLeuGlyAsnAspSerValLysValSerSer 1465

 4111 CTTGCATTCAACCGCTACCGCGCGGGCATTAAGGCAGATTATTCATTCAA 4160
      ::::::   :::         ||||||:::          |||   |||||
 1466 MetSerValLysThrLeuThrAlaGlyLeuAspPheAlaTyrArgPheLy 1482

 4161 ACCGGCGCAACACATTTCCATCACGCCTTATTTGAGCCTGTCCTAT...A 4207
      |   :::   ::::::::::::   |||   ||||||   :::|||   :
 1482 s...ValGlyAsnLeuThrValLysProLeuLeuSerAlaAlaTyrPheA 1498

 4208 CCGATGCCGCTTCCGGCAAAGTCCGAACGCGCGTCAATACCGCCGTATTG 4257
      :::::   :::::::|||   |||::::::      |||:::   :::
 1498 laAsnTyrGlyLysGlyGlyValAsnValGlyGlyAsnSerPheAlaTyr 1514

 4258 GCGCAGGATTTCGGCAAAACCCGCAGTGCGGAATGGGGCGTAAACGCCGA 4307
            |||      :::      ||||||      |||:::
 1515 LysAlaAspAsnGlnGlnTyrSerAla......GlyAlaAlaLeuLe 1529

 4308 AATCAAAGGTTTCACGCTGTCCCTCCACGCTGCCGCCGCCAAGGGGCCGC 4357
         ::::::   ||||||:::::::::::::::   :::||||||   |
 1529 uTyrArgAsnValThrLeuAsnValAsnGlySerIleThrLysGlyLysG 1546

 4358 AATTGGAAGCGCAGCACAGCGCGGGCATCAAATTAGGCTACCGCTGG 4404
      ||||||||   |||   |||:::   ||||||:::      |||:::
 1546 lnLeuGluLysGlnLysSerGlyGlnIleLysIleGlnIleArgPhe 1561

seq_name: pir2:H64106

seq_documentation_block:
IgA-specific metalloendopeptidase (EC 3.4.24.13) type 1 - Haemophilus influenzae (strain
N;Alternate names: immunoglobulin A1 proteinase type 1
C;Species: Haemophilus influenzae
A;Variety: strain Rd KW20
C;Date: 18-Aug-1995 #sequence_revision 18-Aug-1995 #text_change 08-Dec-2000
C;Accession: H64106; A41500
R;Fleischmann, R.D.; Adams, M.D.; White, O.; Clayton, R.A.; Kirkness, E.F.; Kerlavage, A
; Gocayne, J.D.; Scott, J.; Shirley, R.; Liu, L.I.; Glodek, A.; Kelley, J.M.; Weidman, J
, D.M.; Brandon, R.C.; Fine, L.D.; Fritchman, J.L.; Fuhrmann, J.L.; Geoghagen, N.S.M.
Science 269, 496-512, 1995
A;Authors: Gnehm, C.L.; McDonald, L.A.; Small, K.V.; Fraser, C.M.; Smith, H.O.; Vente
A;Title: Whole-genome random sequencing and assembly of Haemophilus influenzae Rd.
A;Reference number: A64000; MUID:95350630
A;Accession: H64106
A;Status: nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-1694 <TIGR>
A;Cross-references: GB:U32779; GB:L42023; NID:g1574009; PIDN:AAC22651.1; PID:g1574019
A;Experimental source: strain Rd KW20
R;Grundy, F.J.; Plaut, A.G.; Wright, A.
Infect. Immun. 58, 320-331, 1990
A;Title: Localization of the cleavage site specificity determinant of Haemophilus inf
A;Reference number: A41500; MUID:90129281
A;Accession: A41500
A;Molecule type: DNA
A;Residues: 1-377 <GRU>
A;Cross-references: GB:X59800
A;Experimental source: strain Rd KW20
A;Note: the authors translated the codon TGG for residue 319 as Thr
C;Function:
A;Description: this proteinase is classified as type 1 because it cleaves at a prolin
C;Superfamily: IgA-specific metalloendopeptidase
C;Keywords: hydrolase; metalloproteinase

alignment_scores:
            Quality: 1587.50              Length:   1859
              Ratio:  1.683                 Gaps:     48
 Percent Similarity:  50.726   Percent Identity: 25.282

alignment_block:
 US-09-303-518D-653 x H64106   ..

 Align seg 1/1  to: H64106  from: 1  to: 1694

   64 CGCTTCTCGCCCGCTTACTTAGCCATATGCCTGTCGTTCGGCATTCTGCC 113
      :::|||:::      ::::::|||:::   ::::::::::::::   ||
    5 LysPheLysLeuAsnPheIleAlaLeuThrValAlaTyrAlaLeuThrPr 21

  114 CCAAGCCCGGGCGGGACACACTTATTTCGGCATCAACTACCAATACTATC 163
      |   ::::::|||:::  :::          ::::::||||||   :::|
   21 oTyrThrGluAlaAlaLeuValArgAspAspValAspTyrGlnIlePheA 38

  164 GCGACTTTGCCGAAAATAAAGGCAAGTTTGCAGTCGGGGCGAAAGATATT 213
      ||||||||||||||||||||||||:::|||::|||||||||   ::::::
   38 rgAspPheAlaGluAsnLysGlyArgPheSerValGlyAlaThrAsnVal 54

  214 GAGGTTTACAACAAAAAAGGGGAGTTGGTCGGCAAATCGATGACGAAAGC 263
      ||||||   :::|||:::::::::   :::|||:::   :::   :::::
   55 GluValArgAspLysAsnAsnHisSerLeuGlyAsnValLeuProAsnGl 71

  264 C...CCGATGATTGATTTTTCTGTGGTATCG...CGTAACGGCGTGGCGG 307
      :   |||||||||||||||||||||||:::      :::   :::|||:
   71 yIleProMetIleAspPheSerValValAspValAspLysArgIleAlaT 88

  308 CATTGGCGGGCGATCAATATATTGTGAGCGTGGCACAT...AACGGCGGC 354
      ::|||   :::   ||||||:::|||:::|||   |||   :::::|||
   88 hrLeuIleAsnProGlnTyrValValGlyValLysHisValSerAsnGly 104

  355 TATAACAATGTTGATTTTGGT..............GCGGAGGGAAGCAA 389
         :::::::::   ||||||                :::::::::||
  105 ValSerGluLeuHisPheGlyAsnLeuAsnGlyAsnMetAsnAsnGlyAs 121

  390 TCCCGATCAGCACCGC.................TTTTCTTACCAAATTG 421
      |       :::||||||                      |||       |
  121 nAlaLysSerHisArgAspValSerSerGluGluAsnArgTyrPheServ 138

  422 TGAAAAGAAATAATTATAAAGCAGGGACTAACGGC.............. 456
      ||:::::::|||:::|||   :::      ||||||
  138 alGluLysAsnGluTyrProThrLysLeuAsnGlyLysAlaValThrThr 154

  457 ............CATCCTTATGGCGGCGATTATCATATGCCGCGTTTGCA 494
```





```
1204 rGluAsnThrThrAspGlnProThrGluArgGluLysThrAlaLysValG 1221

3314 CCACCGAAAAGGCAGAAAGTGTTGCCGAACCGGCCCGGCAGGCAGGC... 3360
1221 luThrGluLysThrGlnGluProProGlnValAlaSerGlnAlaSerPro 1237

3361 .....GGGGA 3365
1238 LysGlnGluGlnSerGluThrValGlnProGlnAlaValLeuGluSerGl 1254

3366 AAATGCC...GGCATTATGCAGGCGGAGGAAGAGAAAAAACGGGTGCAGG 3412
1254 uAsnValProThrValAsnAsnAlaGluGluValGlnAlaGlnLeuGlnT 1271

3413 CGGATAAAGAC..... 3423
1271 hrGlnThrSerAlaThrValSerThrLysGlnProAlaProGluAsnSer 1287

3424 .....ACCGCCTTGGCGAAACAGCGCGAAGCGGAAAC 3455
1288 IleAsnThrGlySerAlaThrAlaIleThrGluThrAlaGluLysSerAs 1304

3456 CCGGCCGGCTACC..... 3468
1304 pLysProGlnThrGluThrAlaAlaSerThrGluAspAlaSerGlnHisL 1321

3469 .....ACCGCC 3474
1321 ysAlaAsnThrValAlaAspAsnSerValAlaAsnAsnSerGluSerSer 1337

3475 TTCCCCCGCGCCCGCCGCGCCCGCCGGGATTTGCCGCAACCGCAG..... 3519
1338 AspProLysSerArgArg...ArgArgSerIleSerGlnProGlnGluTh 1353

3519 ..... 3519
1353 rSerAlaGluGluThrThrAlaAlaSerThrAspGluThrThrIleAlaA 1370

3520 .....CCCCAACCGCAACCCCAACCGCAG..... 3543
1370 spAsnSerLysArgSerLysProAsnArgArgSerArgArgSerValArg 1386

3544 .....CG 3545
1387 SerGluProThrValThrAsnGlySerAspArgSerThrValAlaLeuAr 1403

3546 CGACCTGATCAGCCGTTATGCCAATAGCGGTTTGAGTGAATTTTCCGCC. 3594
1403 gAspLeuThrSerThrAsnThrAsnAlaValIleSerAspAlaMetAlaL 1420

3595 .....ACGCTCAACAGCGTTTTCGCCGTACAGGACGAATTG 3630
1420 ysAlaGlnPheValAlaLeuAsnValGlyLysAlaValSerGlnHisIle 1436

3631 GACCGCGTG...TTTGCCGAAGACCGCCGCAACGCCGTTTGGACAAGCGG 3677
1437 SerGlnLeuGluMetAsnAsnGluGlyGlnTyrAsnValTrpValSerAs 1453

3678 CATCCGGGACACCAAACACTACCGTTCGCAAGATTTCCGCGCCTACCGCC 3727
1453 nThrSerMetAsnGluAsnTyrSerSerSerGlnTyrArgArgPheSerS 1470

3728 AACAAACCGACCTGCGCCAAATCGGTATGCAGAAAAACCTCGGCAGCGGG 3777
1470 erLysSerThrGlnThrGlnLeuGlyTrpAspGlnThrIleSerAsnAsn 1486

3778 ...CGCGTCGGCATCCTGTTTTCGCACAACCGGACCGGAAACACCTTCGA 3824
1487 ValGlnLeuGlyGlyValPheThrTyrValArgAsnSerAsnAsnPheAs 1503

3825 CGACGGCATCGGCAACTCGGCACGGCTTGCCCACGGTGCCGTTTTCGGGC 3874
1503 pLysAlaSerSerLysAsnThrLeuAlaGlnValAsnPheTyrSerLysT 1520

3875 AATACGGCATCGGCAGGTTCGACATCGGCATCAGCGCGGGCGCGGGTTTT 3924
1520 yrTyrAlaAspAsnHisTrpTyrLeuGlyIleAspLeuGlyTyrGlyLys 1536

3925 AGTAGCGGCAGCCTTTCAGACGGCATCAGAGGCAAAATCCGCCGCCGCGT 3974
1537 PheGlnSerAsnLeuLysThrAsnHisAsnAlaLysPheAlaArgHisTh 1553

3975 GCTGCATTACGGCATTCAGGCAAGATACCGCGCAGGTTTCGGCGGATTCG 4024
1553 rAlaGlnPheGlyLeuThrAlaGlyLysAlaPheAsnLeuGlyAsnPheG 1570

4025 GCATCGAACCGCACATCGGCGCAACGCGCTATTTCGTCCAAAAAGCGGAT 4074
1570 lyIleThrProIleValGlyValArgTyrSerTyrLeuSerAsnAlaAsn 1586

4075 TACCGATACGAAAACGTCAATATCGCCACCCCGGGCCTTGCATTCAACCG 4124
1587 PheAlaLeuAlaLysAspArgIleLysValAsnProIleSerValLysTh 1603

4125 CTACCGCGCGGGCATTAAGGCAGATTATTCATTCAAACCGGCGCAACACA 4174
1603 rAlaPheAlaGlnValAspLeuSerTyrThrTyrHisLeuGlyGlu...P 1619

4175 TTTCCATCACGCCTTATTTGAGCCTGTCCTAT...ACCGATGCCGCTTCC 4221
1619 heSerValThrProIleLeuSerAlaArgTyrAspThrAsnGlnGlySer 1635

4222 GGCAAAGTCCGAACGCGCGTCAATACCGCCGTATTGGCGCAGGATTTCGG 4271
1636 GlyLysIle.....AsnValAsnGlnTyrAspPheAlaTyrAsnValGl 1650

4272 CAAAACCCGCAGTGCGGAATGGGGCGTAAACGCCGAAATCAAAGGTTTCA 4321
1650 uAsnGlnGlnGlnTyrAsnAlaGlyLeuLysLeuLysTyrHisAsnValL 1667

4322 CGCTGTCCCTCCACGCTGCCGCCGCCAAGGGGCCGCAATTGGAAGCGCAG 4371
1667 ysLeuSerLeuIleGlyGlyLeuThrLysAlaLysGlnAlaGluLysGln 1683

4372 CACAGCGCGGGCATCAAATTAGGCTAC 4398
1684 LysThrAlaGluLeuLysLeuSerPhe 1692
```

seq_name: pir2:B41859

seq_documentation_block:
IgA-specific metalloendopeptidase (EC 3.4.24.13) type 1 precursor - Haemophilus influ
C;Species: Haemophilus influenzae
A;Variety: strain HK393
C;Date: 04-Mar-1993 #sequence_revision 18-Nov-1994 #text_change 08-Dec-2000
C;Accession: B41859
R;Poulsen, K.; Reinholdt, J.; Kilian, M.
J. Bacteriol. 174, 2913-2921, 1992
A;Title: A comparative genetic study of serologically distinct Haemophilus influenzae
A;Reference number: A41859; MUID:92234949
A;Accession: B41859
A;Status: preliminary; not compared with conceptual translation
A;Molecule type: nucleic acid
A;Residues: 1-1545 <POU>
A;Cross-references: GB:M87490; NID:g148908; PIDN:AAA24967.1; PID:g148909
A;Experimental source: strain HK393
A;Note: sequence extracted from NCBI backbone (NCBIP:97283)
C;Superfamily: IgA-specific metalloendopeptidase
C;Keywords: hydrolase; metalloproteinase

alignment_scores:
Quality: 1587.00 Length: 1723
Ratio: 1.710 Gaps: 47
Percent Similarity: 53.860 Percent Identity: 26.291

```
alignment_block:
 US-09-303-518D-653 x B41859   ..

 Align seg 1/1  to: B41859  from: 1  to: 1545

      64 CGCTTCTCGCCCGCTTACTTAGCCATATGCCTGTCGTTCGGCATTCTGCC 113
         :::|||:::      :::::::|||:::   :::::::::::::   ||
       5 LysPheLysLeuAsnPheIleAlaLeuThrValAlaTyrAlaLeuThrPr 21

     114 CCAAGCCCGGGCGGGACACACTTATTTCGGCATCAACTACCAATACTATC 163
         |   ::::::|||:::   :::         :::::|||||   :::|
      21 oTyrThrGluAlaAlaLeuValArgAspAspValAspTyrGlnIlePheA 38

     164 GCGACTTTGCCGAAAATAAAGGCAAGTTTGCAGTCGGGGCGAAAGATATT 213
         ||||||||||||||||||||||||||||:::|||||||||   ::::::
      38 rgAspPheAlaGluAsnLysGlyLysPheSerValGlyAlaThrAsnVal 54

     214 GAGGTTTACAACAAAAAAGGGGAGTTGGTCGGCAAATCGATGACGAAAGC 263
         ||||||   :::|||:::::::::   :::|||:::  :::   :::::
      55 GluValArgAspLysAsnAsnArgProLeuGlyAsnValLeuProAsnGl 71

     264 C...CCGATGATTGATTTTTCTGTGGTATCG...CGTAACGGCGTGGCGG 307
         :   |||||||||||||||||||||||:::      :::   :::|||:
      71 yIleProMetIleAspPheSerValValAspValAspLysArgIleAlaT 88

     308 CATTGGCGGGCGATCAATATATTGTGAGCGTGGCACAT...AACGGCGGC 354
         ::|||:::::   ||||||:::|||:::|||   |||   ::::::|||
      88 hrLeuValAsnProGlnTyrValValGlyValLysHisValSerAsnGly 104

     355 TATAACAATGTTGATTTTGGT..............GCGGAGGGAAGCAA 389
            :::::::::   ||||||              ::::::::::||
     105 ValSerGluLeuHisPheGlyAsnLeuAsnGlyAsnMetAsnAsnGlyAs 121

     390 TCCCGATCAGCACCGC.................TTTTCTTACCAAATTG 421
         |          ||||||                  |||       |
     121 nAlaLysAlaHisArgAspValSerSerGluGluAsnArgTyrTyrThrV 138

     422 TGAAAAGAAATAATTATAAAGCAGGGACTAACGGC.............. 456
         ||::::::|||:::|||   :::      ||||||
     138 alGluLysAsnGluTyrProThrLysLeuAsnGlyLysAlaValThrThr 154

     457 ............CATCCTTATGGCGGCGATTATCATATGCCGCGTTTGCA 494
                    :::              |||||||:::||||||||||||
     155 GluAspGlnAlaGlnLysArgArgGluAspTyrTyrMetProArgLeuAs 171

     495 CAAATTTGTCACAGATGCAGAACCTGTTGAG...ATGACCAGTTATATGG 541
          ||||||||||||:::::   |||:::|||      |||:::
     171 pLysPheValThrGluValAlaProIleGluAlaSerThrAspSerSerT 188

     542 ATGGGTGGAAATACGCTGATTTAAATAAATACCCTGATCGTGTTCGAATC 591
           :::      |||   :::   :::|||||||||      |||||||:::
     188 hrAlaGlyThrTyrAsnAsnLysAspLysTyrProTyrPheValArgLeu 204

     592 GGAGCAGGCAGACAATAT............................TG 611
         |||:::|||   |||:::                            ||
     205 GlySerGlyThrGlnPheIleTyrGluAsnGlyThrArgTyrGluLeuTr 221

     612 GCGGTCTGATGAAGACGAACCCAATAACCGCGAAAGTTCATATCAT.... 657
         |   :::   |||   :::   ::::::      ::::::|||:::
     221 pLeuGlyLysGluGlyGlnLysSerAsp...AlaGlyGlyTyrAsnLeuL 237

     658 .....ATTGCAAGCGCATATTCTTGGCTCGTCGGTGGCAATACCTTTGCA 702
              ::::::::::||||||::::::   ::::::|||:::   :::
     237 ysLeuValGlyAsnAlaTyrThrTyrGlyIleAlaGlyThrProTyrGlu 253

     703 CAAAATGGATCAGGTGGTGGCACAGTCAACTTAGGTAGC.......... 741
            |||   ::::::   |||   :::::::::|||:::
     254 ValAsnHisGluAsnAspGlyLeuIleGlyPheGlyAsnSerAsnAsnGl 270

     742 ................GAAAAAATTAAACATAGCCCATATGGTTTTTTAC 775
                        |||   ::::::   :::|||          |||
     270 uTyrIleAsnProLysGluIleLeuSerLysLysPro.........LeuT 284

     776 CAACAGGAGGCTCATTTGGCGACAGTGGCTCACCAATGTTTATCTATGAT 825
         :::   :::   :::||||||||||||||||||:::|||:::||||||
     284 hrAsnTyrAlaValLeuGlyAspSerGlySerProLeuPheValTyrAsp 300

     826 GCCCAAAAGCAAAAGTGGTTAATTAATGGGGTATTGCAAACAGGCAACCC 875
           ::::||   ||||||||||:::   |||       :::
     301 ArgGluLysGlyLysTrpLeuPheLeuGlySerTyrAspTyrTrpAlaGl 317

     876 CTATATAGGAAAAAGCAATGGCTTCCAGCTAGTTCGTAAAGATTGGTTCT 925
          |||                                 :::||||:::|||
     317 yTyr.........................AsnLysLysSerTrpGlnG 325

     926 ATGATGAAATCTTTGCTGGAGATACCCATTCAGTATTCTACGAACCACAT 975
               :::|||:::      :::      :::   :::||||||   :::
     325 luTrpAsnIleTyrLysProGluPheAlaGluLysIleTyrGluGlnTyr 341

     976 CAAAATGGGAAATACTTTTTTAACGACAATAATAATGGCGCAGGAAAAAT 1025
                                                 :::||||||:::::
     342 ................................SerAlaGlySerLe 346

    1026 CGATGCCAAACATAAACACTATTCTCTACCTTATAGATTAAAAACACGAA 1075
         :   ::::::         ||||||         :::   ||||||   |
     346 uIleGlySerLysThrAspTyrSerTrpSerSerAsnGlyLysThrSerT 363

    1076 CCGTT............CAATTGTTTAATGTTTCTTTATCCGAGACAGCA 1113
         ||:::           :::   :::|||||:::|||::::::
     363 hrIleThrGlyGlyGluLysSerLeuAsnValAspLeuAlaAspGlyLys 379

    1114 AGAGAACCTGTTTATCATGCTGCAGGTGGGGTCAACAGTTATCGACCCAG 1163
            :::|||
     380 AspLysPro..................................... 382

    1164 ACTGAATAATGGAGAAAATATTTCCTTTATTGACAAAGGAAAAGGTGAAT 1213
            |||:::|||:::::::::::::|||       :::|||:::|||   |
     383 ....AsnHisGlyLysSerValThrPhe......GluGlySerGlyThrL 396

    1214 TGATACTTACCAGCAACATCAACCAAGGCGCGGGCGGTTTGTATTTTGAG 1263
         ||   |||::::::||||||:::||||||||||||||||||:::||||||
     396 euThrLeuAsnAsnAsnIleAspGlnGlyAlaGlyGlyLeuPhePheGlu 412

    1264 GGTAATTTTACGGTC...TCGCCTAAAAACAACGAAACGTGGCAAGGCGC 1310
         |||::::::   |||   :::   ::::::|||   ||||||:::|||||
     413 GlyAspTyrGluValLysGlyThrSerAspAsnThrThrTrpLysGlyAl 429

    1311 GGGCGTTCATATCAGTGATGGCAGTACCGTTACTTGGAAAGTAAACGGCG 1360
         |||||||   :::::::::|||:::||||||||||||||||||:::::
     429 aGlyValSerValAlaGluGlyLysThrValThrTrpLysValHisAsnP 446

    1361 TGGCAAACGACCGCCTGTCCAAAATCGGCAAAGGCACGCTGCTGGTTCAA 1410
                |||||||||:::|||||||||||||||||||||:::|||:::
     446 roGlnTyrAspArgLeuAlaLysIleGlyLysGlyThrLeuIleValGlu 462

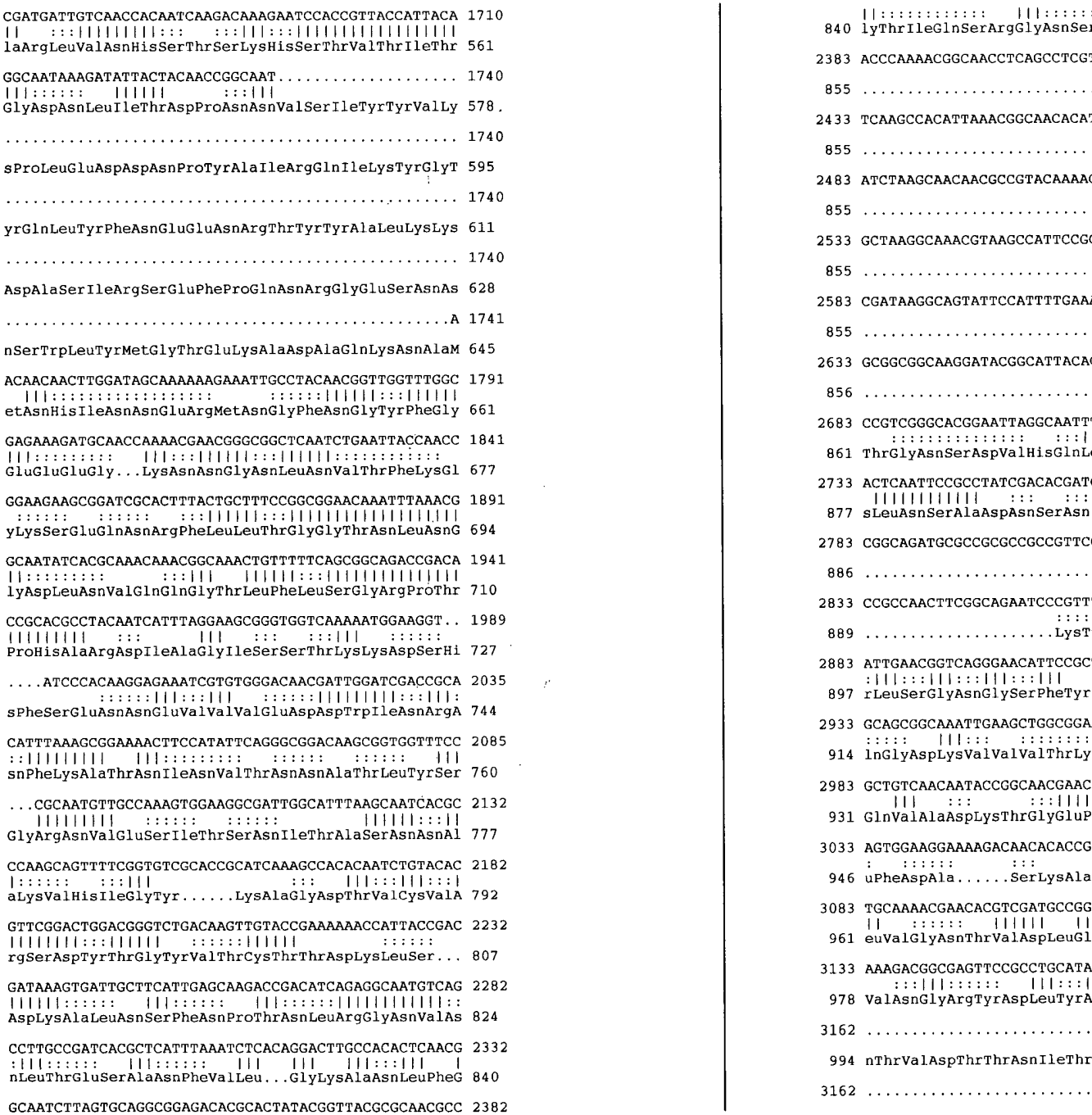
    1411 GCCAAAGGGGAAAACCAAGGCTCGGTCAGCGTGGGCGACGGTAAAGTCAT 1460
         :::   |||:::|||:::||||||:::::|||||||||||   |||||
     463 GlyThrGlyAspAsnLysGlySerLeuLysValGlyAspGlyThrValIl 479

    1461 CTTAGATCAGCAGGCGGACGATCAAGGCAAAAAACAAGCCTTTAGTGAAA 1510
         ||||   ||||||:::::   :::|||   ::::::||||||:::::::
     479 eLeuLysGlnGlnThrAsnGlySerGly...GlnHisAlaPheAlaSerV 495

    1511 TCGGCTTGGTCAGCGGCAGGGGGACGGTGCAACTGAATGCCGATAATCAG 1560
         ::|||:::||||||||||||:::|||:::   ||||||   |||:::|||
     495 alGlyIleValSerGlyArgSerThrLeuValLeuAsnAspAspLysGln 511

    1561 TTCAACCCCGACAAACTCTATTTCGGCTTTCGCGGCGGACGTTTGGATTT 1610
         :::|||::::::::::|||||||||||||||||||||||||||||||||
     512 ValAspProAsnSerIleTyrPheGlyPheArgGlyGlyArgLeuAspLe 528

    1611 GAACGGGCATTCGCTTTCGTTCCACCGCATTCAAAATACCGATGAAGGGG 1660
         |||||||::|||||||:::|||   :::||||:::|||   ||||||||||
     528 uAsnGlyAsnSerLeuThrPheAspHisIleArgAsnIleAspGluGlyA 545
```



```
1011 alProSerValProSerAsnAsnGluGluIleAlaArgValAspGluAla 1027

3162 ..................................................  3162

1028 ProValProProProAlaProAlaThrProSerGluThrThrGluThrVa 1044

3162 ..................................................  3162

1044 lAlaGluAsnSerLysGlnGluSerLysThrValGluLysAsnGluGlnA 1061

3162 ..................................................  3162

1061 spAlaThrGluThrThrAlaGlnAsnArgGluValAlaLysGluAlaLys 1077

3163 ......GTCAAAGAACAAGAGCTTTCCGACAAACTCGGCAAGGCGGGA..  3204
           ||||||   :::      :::::::::::::::::::::|||
1078 SerAsnValLysAlaAsnThrGlnThrAsnGluValAlaGlnSerGlySe 1094

3205 .GAAACAGAGGCCGCCTTGACGGCAAAACAGGCACAACTTGCCGCCAAAC 3253
      ||||||:::   :::   |||::::::      :::   |||:::   :
1094 rGluThrLysGluThrGlnThrThrGluThrLysGluThrAlaThrValG 1111

3254 AACAGGCGGAAAAAGACAACGCGCAAAGC..................... 3282
     :::::   ||||||   ::::::::::::
1111 luLysGluGluLysAlaLysValGluThrGluLysThrGlnGluValPro 1127

3283 ...CTTGACGCGCTGATTGCGGCCGGGCGCAATGCCACCGAAAAGGCAGA 3329
        :::   :::   ::::::      ::::::   :::|||   :::::
1128 LysValThrSerGlnValSerProLysGlnGluGlnSerGluThrValGl 1144

3330 AAGTGTTGCCGAACCGGCCCCGGCAGGCAGGCGGGGAAAATGCCGGCATTA 3379
     :      ||||||||||||||||             ||||||      :
1144 nProGlnAlaGluProAlaArg.............GluAsnAspProThrV 1157

3380 TGCAGGCGGAGGAAGAGAAAAAACGGGTGCAGGCGGATAAAGACACCGCC 3429
     :::::   :::|||   ::::::::::::::::::      ||||||
1157 alAsnIleLysGluProGlnSerGlnThrAsnThrThrAlaAspThrGlu 1173

3430 TTG...GCGAAACAGCGCGAAGCG........................... 3450
          ||||||:::   ::::::
1174 GlnProAlaLysGluThrSerSerAsnValGluGlnProValThrGluSe 1190

3451 .....................................GAAACCC 3457
                                          :::|||
1190 rThrThrValAsnThrGlyAsnSerValValGluAsnProGluAsnThrT 1207

3458 GGCCGGCTACCACCGCCTTCCCC........................... 3480
       ||||||||||||      |||
1207 hrProAlaThrThrGln...ProThrValAsnSerGluSerSerAsnLys 1222

3481 .........CGCGCCCGCCGCGCCCGCCGGGATTTGCCG...CAACCGCA 3518
             |||   ||||||:::   |||:::::|||   :::   ::
1223 ProLysAsnArgHisArgArgSerValArgSerValProHisAsnValGl 1239

3519 GCCCCAACCGCAACCCCAACCGCAGCGC.................GACC 3550
     :|||            :::   :::|||                ||||
1239 uProAlaThrThrSerSerAsnAspArgSerThrValAlaLeuCysAspL 1256

3551 TGATCAGCCGTTATGCCAATAGCGGTTTGAGTGAATTTTCCGCC...... 3594
     ||   |||      :::|||:::   ||||||:::      |||
1256 euThrSerThrAsnThrAsnAlaValLeuSerAspAlaArgAlaLysAla 1272

3595 .........ACGCTCAACAGCGTTTTCGCCGTACAGGACGAATTGGACCG 3635
             :::|||||||         ||||||:::::::::::::::::
1273 GlnPheValAlaLeuAsnValGlyLysAlaValSerGlnHisIleSerGl 1289

3636 CGTG...TTTGCCGAAGACCGCCGCAACGCCGTTTGGACAAGCGGCATCC 3682
     ::::   :::   ::::::   :::      ||||||:::|||:::
1289 nLeuGluMetAsnAsnGluGlyGlnTyrAsnValTrpValSerAsnThrS 1306

3683 GGGACACCAAACACTACCGTTCGCAAGATTTCCGCGCCTACCGCCAACAA 3732
        :::|||:::|||   |||:::::::::|||   :::   ::::::
1306 erMetAsnLysAsnTyrSerSerSerGlnTyrArgArgPheSerSerLys 1322

3733 ACCGACCTGCGCCAAATCGGTATGCAGAAAAACCTCGGCAGCGGG...CG 3779
     :::          |||:::|||   :::::::::::::::::::::   ::
1323 SerThrGlnThrGlnLeuGlyTrpAspGlnThrIleSerAsnAsnValGl 1339

3780 CGTCGGCATCCTGTTTTCGCACAACCGGACCGGAAACACCTTCGACGACG 3829
     :::|||   :::||||::::::   |||::::::|||:::||||||   :
1339 nLeuGlyGlyValPheThrTyrValArgAsnSerAsnAsnPheAspLysA 1356

3830 GCATCGGCAACTCGGCACGGCTTGCCCACGGTGCCGTTTTCGGGCAATAC 3879
     ::   ::::::::::::          :::                |||
1356 laThrSerLysAsnThrLeuAlaGlnValAsnPheTyrSerLysTyrTyr 1372

3880 GGCATCGGCAGGTTCGACATCGGCATCAGCGCGGGCGCGGGTTTTAGTAG 3929
     :::   :::::::::   :::||||||:::   |||   |||     ::
1373 AlaAspAsnHisTrpTyrLeuGlyIleAspLeuGlyTyrGlyLysPheGl 1389

3930 CGGCAGCCTTTCAGACGGCATCAGAGGCAAAATCCGCCGCCGCGTGCTGC 3979
     :::::::|||:::   :::   :::::|||:::   |||::::::   :
1389 nSerLysLeuGlnThrAsnHisAsnAlaLysPheAlaArgHisThrAlaG 1406

3980 ATTACGGCATTCAGGCAAGATACCGCGCAGGTTTCGGCGGATTCGGCATC 4029
     :::::|||:::   |||            :::::::|||:::|||||||||
1406 lnPheGlyLeuThrAlaGlyLysAlaPheAsnLeuGlyAsnPheGlyIle 1422

4030 GAACCGCACATCGGCGCAACGCGCTATTTCGTCCAAAAAGCGGATTACCG 4079
        |||   :::|||:::          :::::::::::::|||||:::
1423 ThrProIleValGlyValArgTyrSerTyrLeuSerAsnAlaAspPheAl 1439

4080 ATACGAAAACGTCAATATCGCCACCCCGGGCCTTGCATTCAACCGCTACC 4129
          :::::::::::::|||   :::      :::::   :::
1439 aLeuAspGlnAlaArgIleLysValAsnProIleSerValLysThrAlaP 1456

4130 GCGCGGGCATTAAGGCAGATTATTCATTCAAACCGGCGCAACACATTTCC 4179
       |||   :::      :::|||::::::      ::::::   :::|||
1456 heAlaGlnValAspLeuSerTyrThrTyrHisLeuGlyGlu...PheSer 1471

4180 ATCACGCCTTATTTGAGCCTGTCCTATACCGATGCCGCTTCCGGCAAAGT 4229
     :::||||||   ||||||      |||   ||||||   :::|||:::
1472 ValThrProIleLeuSerAlaArgTyr...AspAlaAsnGlnGlySerGl 1487

4230 CCGAACGCGCGTCAATACCGCCGTATTGGCGCAGGATTTCGGCAAAACCC 4279
      :::   :::||||||         :::|||   :::      :::   :
1487 yLysIleAsnValAsnGlyTyrAspPheAlaTyrAsnValGluAsnGlnG 1504

4280 GCAGTGCGGAATGGGGCGTAAACGCCGAAATCAAAGGTTTCACGCTGTCC 4329
     :::::   :::   |||::::::   :::      :::      ||||||
1504 lnGlnTyrAsnAlaGlyLeuLysLeuLysTyrHisAsnValLysLeuSer 1520

4330 CTCCACGCTGCCGCCGCCAAGGGGCCGCAATTGGAAGCGCAGCACAGCGC 4379
     |||   ::::::   :::|||:::   |||   |||   |||   :::||
1521 LeuIleGlyGlyLeuThrLysAlaLysGlnAlaGluLysGlnLysThrAl 1537

4380 GGGCATCAAATTAGGCTAC 4398
     |   :::||||||::::::
1537 aGluLeuLysLeuSerPhe 1543
```

seq_name: pir2:A81018

seq_documentation_block:
serine-type peptidase NMB1998 [imported] - Neisseria meningitidis (strain MC58 serogr
C;Species: Neisseria meningitidis
C;Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 19-Jan-2001
C;Accession: A81018
R;Tettelin, H.; Saunders, N.J.; Heidelberg, J.; Jeffries, A.C.; Nelson, K.E.; Eisen,
Hickey, E.K.; Haft, D.H.; Salzberg, S.L.; White, O.; Fleischmann, R.D.; Dougherty, B.
ri, H.; Qin, H.; Vamathevan, J.; Gill, J.; Scarlato, V.; Masignani, V.; Pizza, M.
Science 287, 1809-1815, 2000
A;Authors: Grandi, G.; Sun, L.; Smith, H.O.; Fraser, C.M.; Moxon, E.R.; Rappuoli, R.;
A;Title: Complete genome sequence of Neisseria meningitidis serogroup B strain MC58.
A;Reference number: A81000; MUID:20175755
A;Accession: A81018
A;Status: preliminary

```
A;Molecule type: DNA
A;Residues: 1-1431 <TET>
A;Cross-references: GB:AE002549; GB:AE002098; NID:g7227258; PIDN:AAF42325.1; PID:g722726
A;Experimental source: serogroup B, strain MC58
C;Genetics:
A;Gene: NMB1998


alignment_scores:
            Quality: 1585.50              Length:   1668
              Ratio:  1.829                Gaps:     50
 Percent Similarity:  51.978    Percent Identity: 28.118

alignment_block:
 US-09-303-518D-653 x A81018   ..

 Align seg 1/1  to: A81018  from: 1  to: 1431

      55 GGCCGCATCCGCTTCTCGCCCGCTTACTTAGCCATATGCCTGTCGTTCGG 104
         |||   :::      ||||||||||||   :::|||      :::
      17 GlyLeuPheAlaValSerProAlaTyrSerSerIleValArgAsn..... 31

     105 CATTCTGCCCCAAGCCCGGGCGGGACACACTTATTTCGGCATCAACTACC 154
                                                 ::::::||||
      32 .....AspValAspTyrG 36

     155 AATACTATCGCGACTTTGCCGAAAATAAAGGCAAGTTTGCAGTCGGGGCG 204
         |||||:::||||||||||||||||||||||||   |||:::|||||||||
      36 lnTyrPheArgAspPheAlaGluAsnLysGlyAlaPheThrValGlyAla 52

     205 AAAGATATTGAGGTTTACAACAAAAAAGGGGAGTTGGTCGGCAAATCGAT 254
         ::::::|||::::::   :::|||:::|||:::::::::|||:::   ::
      53 SerAsnIleSerIleGlnAspLysGlnGlyLysIleLeuGlyArgValLe 69

     255 GACGAAAGCCCCGATGATTGATTTTTCTGTGGTATCGCGT...AACGGCG 301
         ::::      ||||||   ||||||   |||   :::|||   :::::::
      69 uAsnGlyIleProMetProAspPheArgValSerAsnArgGlnThrAlaI 86

     302 TGGCGGCATTGGCGGGCGATCAATATATTGTGAGCGTGGCACATAACGGC 351
         ::|||:::|||:::      ||||||:::   ||||||   ||||||
      86 leAlaThrLeuValHisProGlnTyrValAsnSerValLysHisAsnVal 102

     352 GGCTATAACAATGTTGATTTTGGTGCGGAGGGAAGCAATCCCGATCAGCA 401
         ||||||:::::::::::::||||||   :::   :::||||||:::::::
     103 GlyTyrGlySerIleGlnPheGlyAsnAspThrGlnAsnProGluGluGl 119

     402 CCGCTTTTCTTACCAAATTGTGAAAAGAAATAATTATAAAGCAGGGACTA 451
         :   ::::::|||::::::|||:::|||                    |
     119 nAlaTyrThrTyrArgLeuValSerArg.....................A 129

     452 ACGGCCATCCTTATGGCGGCGATTATCATATGCCGCGTTTGCACAAATTT 501
         ||   ||||||         ||||||||||:::||||||||||:::|||:::
     129 snProHisPro...AspTyrAspTyrHisLeuProArgLeuAsnLysLeu 144

     502 GTCACAGATGCAGAACCTGTTGAGATGACCAGT..................... 534
         ||||||:::   :::|||:::   ::::::|||
     145 ValThrGluIleSerProThrAlaLeuSerSerValProLeuLeuGlyAs 161

     535 ....TATATGGATGGGTGGAAATACGCTGATTTAAATAAATACCCTGATC 580
                 :::      |||   |||   ::::::::::|||
     161 nGlyGlnProLysAlaAsnAlaTyrLeuAspThrAspArgProProTyrP 178

     581 GTGTTCGAATCGGAGCAGGCAGACAATATTGGCGGTCTGATGAAGACGAA 630
           ||||||:::|||:::|||   |||      |||:::   :::
     178 heValArgLeuGlySerGlyThrGlnGlnValArgLysAlaAspGly... 193

     631 CCCAATAACCGCGAAAGTTCATATCATATTGCAAGCGCATATTCTTGGCT 680
                       :::      :::   |||   ||||||::::::||
     194 ..............ThrArgThrArgThrAlaProAlaTyrGlnTyrLe 205

     681 CGTCGGTGGCAATACCTTTGCACAAAATGGATCAGGTGGTGGCACAGTCA 730
         |:::||||||:::   :::         |||      :::      :::
     205 uThrGlyGlyThrProLeuLysValLeuGlyPheGlnAsnHisGlyLeuL 222

     731 ACTTAGGTAGCGAAAAAATTAAACATAGCCCATATGGTTTTTTACCAACA 780
           :::|||:::   ::::::      :::|||         |||   |||
     222 euValGlyGly...SerLeuThrAspGlnPro.....LeuAsnThr 234

     781 GGAGGCTCATTTGGCGACAGTGGCTCACCAATGTTTATCTATGATGCCCA 830
           :::      ||||||||||||||||||:::|||   :::|||   ::
     235 TyrAlaIleAlaGlyAspSerGlySerProLeuPheAlaPheAspLysHi 251

     831 AAAGCAAAAGTGGTTAATTAATGGGGTATTGCAAACAGGCAACCCCTATA 880
         ::::::::::|||:::::   ||||||||||:::|||         |||
     251 sGluAsnArgTrpValLeuAlaGlyValLeuSerThr.....TyrA 265

     881 TAGGAAAAAGCAATGGCTTCCAG......CTAGTTCGTAAAGATTGGTTC 924
           |||   :::|||   |||:::      :::|||   :::      |||
     265 laGlyPheAspAsnPhePheAsnLysTyrIleValThrGlnProGluPhe 281

     925 TATGATGAAATCTTTGCTGGAGATACCCATTCAGTATTCTACGAACCACA 974
               :::   :::                          ||||||   ::
     282 IleArgSerThrIleArgGln.................TyrGluThrAr 292

     975 T..............CAAAATGGGAAATACTTTTTTAACGACAATAATA 1009
         :                :::::::::   :::::::::||||||:::|
     292 gLeuAspValGlyLeuThrThrAsnGluLeuIleTrpArgAspAsnGlyA 309

    1010 ATGGCGCAGGAAAAATCGATGCCAAACATAAACACTATTCTCTACCTTAT 1059
         |||||   :::   :::::::::   :::::::::   :::||||||
     309 snGlyAsnSerThrLeuGlnGlyLeuAsnGluArgIleThrLeuPro... 324

    1060 AGATTAAAAACACGAACCGTTCAATTGTTTAATGTTTCTTTATCCGAGAC 1109
                                  :::   |||   |||||:::
     325 ........................IleAlaAsnProSerLeuAla..... 331

    1110 AGCAAGAGAACCTGTTTATCATGCTGCAGGTGGGGTCAACAGTTATCGAC 1159
                                                        |
     332 ..............................................P 332

    1160 CCAGACTGAATAATGGAGAAAATATTTCCTTTATTGACAAAGGAAAA... 1206
         ||:::   |||:::::::::::::::::          |||   ||||||
     332 roGln...AsnAspSerArgHisMetProSerGluAspAlaGlyLysThr 347

    1207 ............................GGTGAATTGATACTTACCAGCAA 1229
                                                |||:::|||::::::||
     348 LeuIleLeuSerSerArgPheAspAsnLysThrLeuMetLeuAlaAspAs 364

    1230 CATCAACCAAGGCGCGGGCGGTTTGTATTTTGAGGGTAATTTTACGGTCT 1279
         ||||||||||||||||||:::|||   |||:::::::::::::::::::
     364 nIleAsnGlnGlyAlaGlyAlaLeuGlnPheAspSerAsnPheThrValV 381

    1280 CGCCTAAAAACAACGAAACGTGGCAAGGCGCGGGCGTTCATATCAGTGAT 1329
              ||||||:::   ||||||||||||||||||||   :::::|||
     381 alGlyLysAsnHis...ThrTrpGlnGlyAlaGlyValIleValAlaAsp 396

    1330 GGCAGTACCGTTACTTGGAAAGTAAACGGCGTGGCAAACGACCGCCTGTC 1379
         |||:::   |||   |||:::|||::::::      :::||||||||||
     397 GlyLysArgValPheTrpGlnValSerAsnProLysGlyAspArgLeuSe 413

    1380 CAAAATCGGCAAAGGCACGCTGCTGGTTCAAGCCAAAGGGGAAAACCAAG 1429
         ||||:::|||   |||||||||:::::::::::::::||   |||||||
     413 rLysLeuGlyAlaGlyThrLeuIleAlaAsnGlyGlnGlyIleAsnGlnG 430

    1430 GCTCGGTCAGCGTGGGCGACGGTAAAGTCATCTTAGATCAGCAGGCGGAC 1479
         ||:::::::|||:::|||:::|||   |||:::|||   |||:::|||
     430 lyAspIleSerIleGlyGluGlyThrValValLeuAlaGlnLysAlaAla 446

    1480 GATCAAGGCAAAAAACAAGCCTTTAGTGAAATCGGCTTGGTCAGCGGCAG 1529
         ::::::|||:::||||||||||:::::::::|||::::::|||||||||
     447 SerAspGlySerLysGlnAlaPheAsnGlnValGlyIleThrSerGlyAr 463

    1530 GGGGACGGTGCAACTGAATGCCGATAATCAGTTCAACCCCGACAAACTCT 1579
         |||||||:::   |||      ::::::|||:::::|||::::::||||
     463 gGlyThrAlaValLeuAlaAspSerGlnGlnIleLysProGluAsnLeuT 480
```



```
 937 erHisGlnAlaArgPheThrLeuGlnAsnGlyTyrAlaAspLeuGlyAla 953

3112 TGGCGTTATCAGCTTATCCGCAAAGACGGCGAGTTCCGCCTGCATAATCC 3161
     :::||||||   |||   ::::::::::::   :::   |||:::|||||
 954 TyrArgTyrIleLeuArgLysAsnAsnAsnGlyTyrSerLeuTyrAsnPr 970

3162 GGTCAAAGAACAAGAGCTT.............................. 3180
     |:::||||||   ||||||
 970 oLeuLysGluAlaGluLeuGlnIleGluAlaThrArgAlaGluHisGluA 987

3180 ............................................... 3180

 987 rgAsnGlnGlnAlaTyrAsnGlnLeuGlnAlaThrAspIleSerArgGln 1003

3181 ............TCCGACAAACTCGGCAAGGCG................. 3201
                 ||||||         :::|||
1004 ValGlnHisAspSerAspAlaThrArgGlnAlaLeuGlnAlaTrpGlnAs 1020

3202 .GGAGAAACAGAGGCCGCCTTGACGGCAAAACAGGCACAACTTGCCGCCA 3250
      ::::::||||||   |||         :::|||:::|||      :::
1020 nSerGlnThrGluLeuAlaArgIleAspSerGlnValGlnTyrLeuSerA 1037

3251 AACAACAGGCGGAAAAAGAC.....................AACGCGCAA 3279
       |||      ::   |||                       :::||||||
1037 laGlnLeuLysGlnThrAspProLeuThrGlyIleLeuThrArgAlaGln 1053

3280 AGCCTTGACGCGCTGATTGCGGCCGGGCGCAATGCC.............. 3315
     :::|||   |||       |||   |||   :::|||
1054 AsnLeuCysAla......AlaGlnGlyTyrSerAlaAspIleCysArgGl 1068

3315 ............................................... 3315

1068 nValAlaLysAlaAlaAspThrAsnAspLeuThrLeuPheGluThrGluL 1085

3316 .....................ACCGAAAAGGCAGAAAGTGTTGCCGAACCG 3345
                          :::|||   |||||||||       :::
1085 euAspThrTyrIleGluArgValGluMetAlaGluSerGluLeuAspLys 1101

3346 GCCCGGCAGGCAGGCGGGGAAAATGCCGGCATTATGCAGGCGGAGGAAGA 3395
     |||||||||:::|||                        ||||||   |||
1102 AlaArgGlnGlyGlyAsp..............AlaGlnAlaValGluTh 1113

3396 GAAAAAACGGGTGCAGGCGGATAAAGACACCGCCTTGGCGAAACAG...C 3442
       :::::::::       :::       :::   |||:::::|||   :
1113 rAlaArgHisAlaTyrLeuAsnAlaLeuAsnArgLeuSerArgGlnIleH 1130

3443 GCGAAGCGGAAACCCGGCCGGCTACCACCGCCTTCCCCCGCGCCCGCCGC 3492
     :::::   :::|||      |||         :::|||:::
1130 isSerLeuLysThrGlyValAlaGlyIleArgMetProAsnLeu...... 1144

3493 GCCCGCCGGGATTTGCCGCAACCGCAGCCCCAACCGCAACCCCAACCGCA 3542

1144 ............................................... 1144

3543 GCGCGACCTGATCAGCCGTTATGCCAATAGCGGTTTGAGTGAATTTTCCG 3592
       :::||||||||||||   ||||||:::::::::||||||   :::|
1145 .AlaGluLeuIleSerArgSerAlaAsnThrAlaValSerGluGlnAlaA 1161

3593 CCACGCTCAACAGCGTTTTCGCCGTACAGGACGAATTGGACCGCGTGTTT 3642
     ||      ::::::         :::      ::::::||||||   :::
1161 laTyrAsnThrGlyArgGlnGlnAlaGlyArgArgIleAspArgHisLeu 1177

3643 GCCGAAGACCGCCGCAACGCCGTTTGGACAAGCGGCATCCGGGACACCAA 3692
     ::::::   :::::::::   :::|||   :::         :::
1178 ThrAspProGlnGlnGlnAsnIleTrpLeuGluThrGlyThrGlnGlnTh 1194

3693 ACACTACCGTTCGCAAGATTTCCGCGCCTACCGCCAACAAACCGACCTGC 3742
       |||:::|||       |||   |||:::|||   |||:::
1194 rAspTyrHisSerGlyThrHisArgProTyrGlnGlnThrThrAsnTyrA 1211

3743 GCCAAATCGGTATGCAGAAAAACCTCGGCAGCGGG...CGCGTCGGCATC 3789
       :::||||||:::|||   ::::::   :::         ||||||
1211 laHisIleGlyIleGlnThrGlyIleThrAspArgLeuSerValGlyThr 1227
```

```
3790 CTGTTTTCGCACAACCGGACCGGAAACACCTTCGACGACGGCATCGGC.. 3837
     :::::::::   :::||||||:::|||   ||||||:::|||:::::::
1228 IleLeuThrAspGluArgThrAsnAsnArgPheAspGluGlyValSerAl 1244

3838 ....AACTCGGCACGGCTTGCCCACGGTGCCGTTTTCGGGCAATACGGCA 3883
         |||   ::::::   ||||||      |||   |||:::   |||
1244 aArgAsnArgSerAsnGlyAlaHisLeuPheValLysGlyGluAsnGlyA 1261

3884 TCGGCAGGTTCGACATCGGCATCAGCGCGGGCGCGGGTTTTAGTAGCGGC 3933
                    :::      :::|||      |||:::|||:::::
1261 la............LeuPheAlaAlaAlaAspLeuGlyTyrSerAsnSer 1273

3934 ......AGCCTTTCAGACGGCATCAGAGGCAAAATCCGCCGCCGCCGTGCT 3977
            :::::::|||         :::   :::||||||:::::::
1274 ArgThrArgPheThrAspTyrAspGlyAlaAlaValArgArgHisAlaTr 1290

3978 GCATTACGGCATTCAGGCAAGATACCGCGCAGGTTTCGGCGGATTCGGCA 4027
           |||||::::::      :::            |||:::::::
1290 pAspAlaGlyIleAsnThrGlyIleLysIleAsp...ThrGlyIleAsnL 1306

4028 TCGAACCGCACATCGGC.....................GCAACGCGCTAT 4056
     :::::|||:::   |||                     ::::::||||||
1306 euArgProTyrAlaGlyIleArgIleAsnArgSerAsnGlyAsnArgTyr 1322

4057 TTCGTCCAAAAAGCGGATTACCGATACGAAAACGTCAATATCGCCACCCC 4106
        ::::::   |||::                       |||   :::||
1323 ValLeuAspGlyAlaGlu.....................IleAsnSerPr 1332

4107 GGGCCTTGCATTCAACCGCTACCGCGCGGGCATTAAGGCAGATTATTCAT 4156
     |:::         :::   ::::::||||||||::   |||   :::
1332 oAlaGlnIleGlnThrThrTrpHisAlaGlyIleArgLeuAspLysThrV 1349

4157 TCAAACCGGCGCAACACATTTCCATCACGCCTTATTTGAGCCTGTCCTAT 4206
       :::   :::|||      ::::::||||||   :::|||   :::|||
1349 alGluLeuGlyGln...AlaLysLeuThrProAlaPheSerSerAspTyr 1364

4207 ......ACCGATGCCGCTTCCGGCAAAGTCCGAACGCGCGTCAATACCGC 4250
           |||         ||||||::::::   :::   ||||||
1365 TyrHisThrArgGlnAsnSerGlySerAlaLeuSer...ValAsnAspAr 1380

4251 CGTATTGGCGCAGGATTTCGGCAAAACCCGCAGTGCGGAATGGGGCGTAA 4300
      :::|||   |||:::   :::                       ::::
1380 gThrLeuLeuGlnGlnAlaAlaHisGlyThrLeuHisThrLeuGlnIleA 1397

4301 ACGCCGAAATCAAAGGTTTCACGCTGTCCCTCCACGCTGCCGCCGCCAAG 4350
     ::|||      ||||||::::::   :::||||||||||||   :::|||
1397 spAlaGlyTyrLysGlyTrpAsnAlaLysLeuHisAlaAlaTyrGlyLys 1413

4351 GGGCCGCAATTGGAAGCGCAGCACAGCGCGGGCATCAAATTAGGCTACCG 4400
           :::         :::   :::||||||||||||||:::||||||::
1414 AspSerAsnThrAlaArgHisLysGlnAlaGlyIleLysIleGlyTyrAs 1430

4401 CTGG 4404
     :|||
1430 nTrp 1431
```

seq_name: pir2:A41859

seq_documentation_block:
IgA-specific metalloendopeptidase (EC 3.4.24.13) type 1 precursor - Haemophilus influ
C;Species: Haemophilus influenzae
A;Variety: strain HK715
C;Date: 04-Mar-1993 #sequence_revision 18-Nov-1994 #text_change 08-Dec-2000
C;Accession: A41859
R;Poulsen, K.; Reinholdt, J.; Kilian, M.
J. Bacteriol. 174, 2913-2921, 1992
A;Title: A comparative genetic study of serologically distinct Haemophilus influenzae
A;Reference number: A41859; MUID:92234949
A;Accession: A41859
A;Status: preliminary; not compared with conceptual translation
A;Molecule type: nucleic acid
A;Residues: 1-1702 <POU>

```
A;Cross-references: GB:M87489; NID:g148906; PIDN:AAA24966.1; PID:g148907
A;Experimental source: strain HK715
A;Note: sequence extracted from NCBI backbone (NCBIP:97282)
C;Superfamily: IgA-specific metalloendopeptidase
C;Keywords: hydrolase; metalloproteinase


alignment_scores:
           Quality: 1583.50              Length:   1867
             Ratio:  1.679                Gaps:     48
 Percent Similarity:  50.509   Percent Identity: 25.174

alignment_block:
 US-09-303-518D-653 x A41859   ..

 Align seg 1/1  to: A41859  from: 1  to: 1702

      64 CGCTTCTCGCCCGCTTACTTAGCCATATGCCTGTCGTTCGGCATTCTGCC 113
         :::|||:::      :::::::|||:::   :::::::::::::::   ||
       5 LysPheLysLeuAsnPheIleAlaLeuThrValAlaTyrAlaLeuThrPr 21

     114 CCAAGCCCGGGCGGGACACACTTATTTCGGCATCAACTACCAATACTATC 163
         |   ::::::|||:::   :::         ::::::||||||   :::|
      21 oTyrThrGluAlaAlaLeuValArgAspAspValAspTyrGlnIlePheA 38

     164 GCGACTTTGCCGAAAATAAAGGCAAGTTTGCAGTCGGGGCGAAAGATATT 213
         ||||||||||||||||||||||||:::|||:::|||||||||   ::::::
      38 rgAspPheAlaGluAsnLysGlyArgPheSerValGlyAlaThrAsnVal 54

     214 GAGGTTTACAACAAAAAAGGGGAGTTGGTCGGCAAATCGATGACGAAAGC 263
         ||||||   :::|||:::::::::   :::|||:::   :::   :::::
      55 GluValArgAspLysAsnAsnHisSerLeuGlyAsnValLeuProAsnGl 71

     264 C...CCGATGATTGATTTTTCTGTGGTATCG...CGTAACGGCGTGGCGG 307
         :   |||||||||||||||||||||||:::      :::   :::|||:
      71 yIleProMetIleAspPheSerValValAspValAspLysArgIleAlaT 88

     308 CATTGGCGGGCGATCAATATATTGTGAGCGTGGCACAT...AACGGCGGC 354
         ::|||   :::   |||||||:::|||:::|||   |||   ::::::|||
      88 hrLeuIleAsnProGlnTyrValValGlyValLysHisValSerAsnGly 104

     355 TATAACAATGTTGATTTTGGT..............GCGGAGGGAAGCAA 389
         :::::::::   ||||||                 ::::::::::||
     105 ValSerGluLeuHisPheGlyAsnLeuAsnGlyAsnMetAsnAsnGlyAs 121

     390 TCCCGATCAGCACCGC.................TTTTCTTACCAAATTG 421
         |      :::||||||                      |||      |
     121 nAspLysSerHisArgAspValSerSerGluGluAsnArgTyrPheServ 138

     422 TGAAAAGAAATAATTATAAAGCAGGGACTAACGGC.............. 456
         ||:::::::|||:::|||   :::      ||||||
     138 alGluLysAsnGluTyrProThrLysLeuAsnGlyLysAlaValThrThr 154

     457 ............CATCCTTATGGCGGCGATTATCATATGCCGCGTTTGCA 494
                     :::            ||||||:::||||||||||||||
     155 GluAspGlnThrGlnLysArgArgGluAspTyrTyrMetProArgLeuAs 171

     495 CAAATTTGTCACAGATGCAGAACCTGTTGAG...ATGACCAGTTATATGG 541
          ||||||||||||::::::   |||:::|||      |||:::      |
     171 pLysPheValThrGluValAlaProIleGluAlaSerThrAlaSerSerA 188

     542 ATGGGTGGAAATACGCTGATTTAAATAAATACCCTGATCGTGTTCGAATC 591
         ||:::      |||   |||   ||||||||||||      ||||||:::
     188 spAlaGlyThrTyrAsnAspGlnAsnLysTyrProAlaPheValArgLeu 204

     592 GGAGCAGGCAGACAATATTGGCGGTCTGATGAAGACGAA.......... 630
         |||:::|||   |||:::      :::      |||:::
     205 GlySerGlySerGlnPheIleTyrLysLysGlyAspAsnTyrSerLeuIl 221

     631 .CCCAATAACCGCGAAAGTTCATATCAT............ATTGCAAGCG 667
          ||||||:::|||   :::   :::           :::::::::|
     221 eLeuAsnAsnHisGluValGlyGlyAsnAsnLeuLysLeuValGlyAspA 238

     668 CATATTCTTGGCTCGTCGGTGGCAATACCTTTGCACAAAATGGATCAGGT 717
         |||||::::::   :::::::|||:::   :::      |||   ::::::
     238 laTyrThrTyrGlyIleAlaGlyThrProTyrLysValAsnHisGluAsn 254

     718 GGTGGCACAGTCAACTTAGGTAGCGAAAAAATTAAACATAGC...CCATA 764
         :::|||   ::::::::::|||:::::|||   :::||||||   |||
     255 AsnGlyLeuIleGlyPheGlyAsnSerLysGluGluHisSerAspProLy 271

     765 TGGTTTTTTACCAACAGGA.............GGCTCATTTGGCGACA 799
          |||:::|||                     :::   :::|||||||
     271 sGlyIleLeuSerGlnAspProLeuThrAsnTyrAlaValLeuGlyAspS 288

     800 GTGGCTCACCAATGTTTATCTATGATGCCCAAAAGCAAAAGTGGTTAATT 849
         ||||||||||||:::|||:::||||||   :::|||   |||||||||:::
     288 erGlySerProLeuPheValTyrAspArgGluLysGlyLysTrpLeuPhe 304

     850 AATGGGGTATTGCAAACAGGCAACCCCTATATAGGAAAAAGCAATGGCTT 899
            |||      :::            |||      ||||||
     305 LeuGlySerTyrAspPheTrpAlaGlyTyrAsnLysLysSerTrpGlnGl 321

     900 CCAGCTAGTTCGTAAAGATTGGTTCTATGATGAAATCTTTGCTGGAGATA 949
               :::   |||      |||         ::::::      ||||
     321 uTrpAsnIleTyrLysProGluPheAlaLysThrValLeuAspLysAspT 338

     950 CCCATTCAGTATTCTACGAACCACATCAAAATGGGAAATACTTTTTTAAC 999
         ||
     338 hr.................................................. 338

    1000 GACAATAATAATGGCGCAGGAAAAATCGATGCCAAACATAAACACTATTC 1049
                        ||||||::::::   :::::::::   :::|||::
     339 ..............AlaGlySerLeuThrGlySerAsnThrGlnTyrAs 350

    1050 TCTACCTTATAGATTAAAAACACGAACCGTT............CAATTGT 1087
         :                ||||||   ::::::          :::   :
     350 nTrpAsnProThrGlyLysThrSerValIleSerAsnGlySerGluSerL 367

    1088 TTAATGTTTCTTTATCCGAGACAGCAAGAGAACCTGTTTATCATGCTGCA 1137
         ::|||||:::|||   ::::::::::::::
     367 euAsnValAspLeuPheAspSerSerGlnAsp.................. 377

    1138 GGTGGGGTCAACAGTTATCGACCCAGACTGAATAATGGAGAAAATATTTC 1187
                    :::          :::   |||:::|||::::::::::::
     378 ............ThrAspSerLysLysAsnAsnHisGlyLysSerValTh 390

    1188 CTTTATTGACAAAGGAAAAGGTGAATTGATACTTACCAGCAACATCAACC 1237
         ::::       :::|||:::|||   |||   |||:::::::||||||:::|
     390 rLeu......ArgGlySerGlyThrLeuThrLeuAsnAsnAsnIleAspG 405

    1238 AAGGCGCGGGCGGTTTGTATTTTGAGGGTAATTTTACGGTC...TCGCCT 1284
         ||||||||||||||||:::||||||||::::::   |||   :::
     405 lnGlyAlaGlyGlyLeuPhePheGluGlyAspTyrGluValLysGlyThr 421

    1285 AAAAACAACGAAACGTGGCAAGGCGCGGGCGTTCATATCAGTGATGGCAG 1334
         :::::::::   ||||||:::||||||||||||   ::::::||||||::
     422 SerAspSerThrThrTrpLysGlyAlaGlyValSerValAlaAspGlyLy 438

    1335 TACCGTTACTTGGAAAGTAAACGGCGTGGCAAACGACCGCCTGTCCAAAA 1384
         :||||||||||||||||||::::::      :::||||||||||:::||||
     438 sThrValThrTrpLysValHisAsnProLysSerAspArgLeuAlaLysI 455

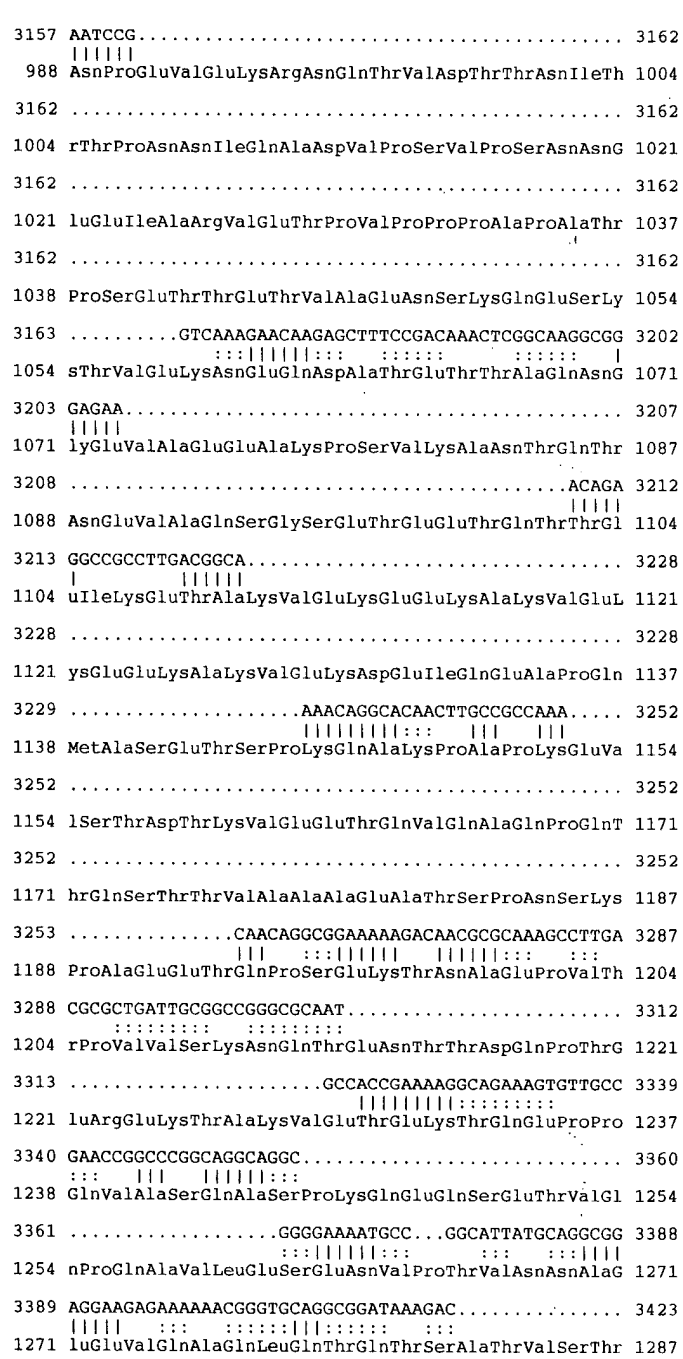
    1385 TCGGCAAAGGCACGCTGCTGGTTCAAGCCAAAGGGGAAAACCAAGGCTCG 1434
         ||||||||||||||||||:::|||::::::||||||||||||:::||||||
     455 leGlyLysGlyThrLeuIleValGluGlyLysGlyGluAsnLysGlySer 471

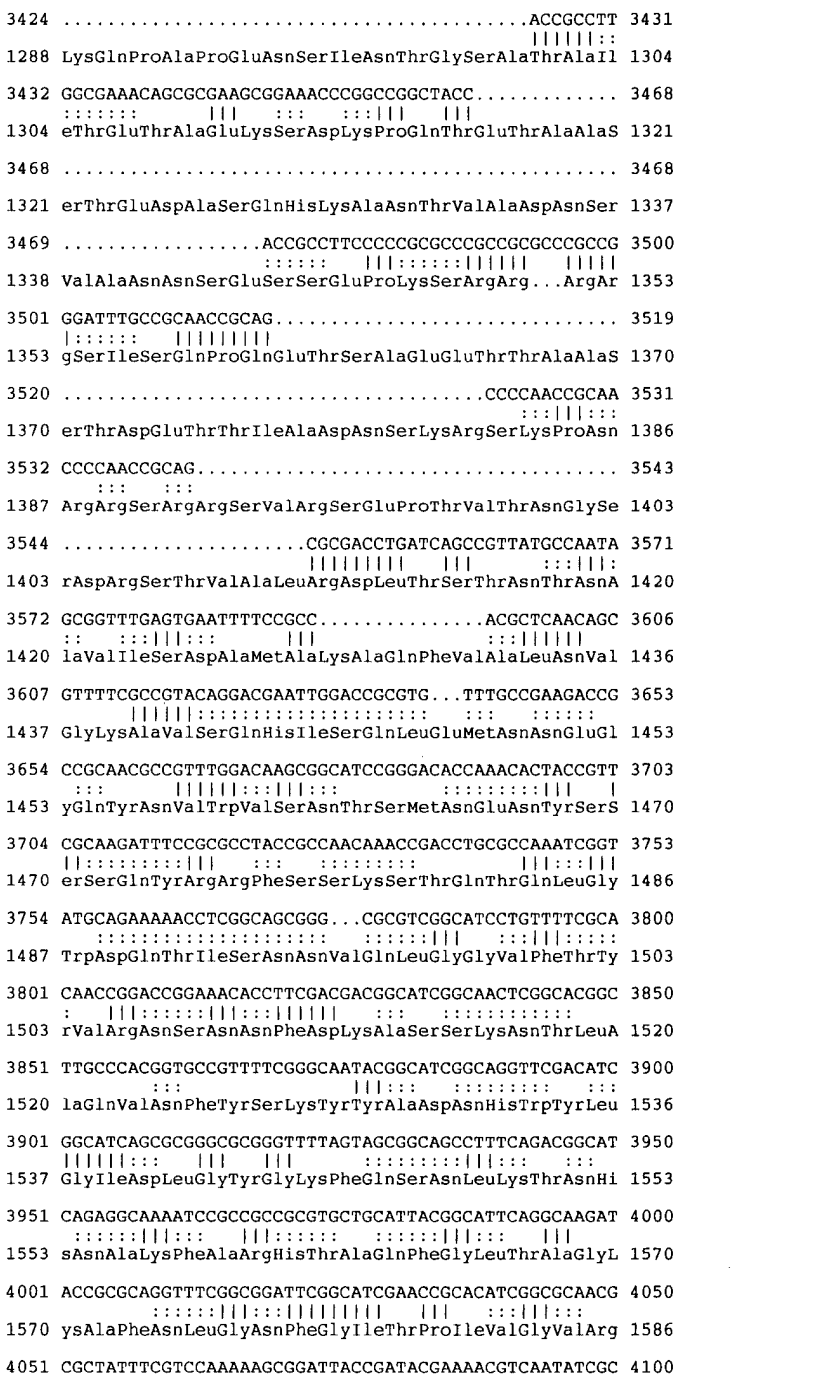
    1435 GTCAGCGTGGGCGACGGTAAAGTCATCTTAGATCAGCAGGCGGACGATCA 1484
         ::::::||||||||||||   ||||||||   ||||||||||||   ::
     472 LeuLysValGlyAspGlyThrValIleLeuLysGlnGlnAlaAspAlaAs 488

    1485 AGGCAAAAAACAAGCCTTTAGTGAAATCGGCTTGGTCAGCGGCAGGGGGA 1534
         ::::|||   :::|||||||||:::::::|||:::||||||||||||:::|
     488 nAsnLysValLysAlaPheSerGlnValGlyIleValSerGlyArgSerT 505

    1535 CGGTGCAACTGAATGCCGATAATCAGTTCAACCCCGACAAACTCTATTTC 1584
```




3157 AATCCG 3162
988 AsnProGluValGluLysArgAsnGlnThrValAspThrThrAsnIleTh 1004
3162 3162
1004 rThrProAsnAsnIleGlnAlaAspValProSerValProSerAsnAsnG 1021
3162 3162
1021 luGluIleAlaArgValGluThrProValProProProAlaProAlaThr 1037
3162 3162
1038 ProSerGluThrThrGluThrValAlaGluAsnSerLysGlnGluSerLy 1054
3163 .....GTCAAAGAACAAGAGCTTTCCGACAAACTCGGCAAGGCGG 3202
1054 sThrValGluLysAsnGluGlnAspAlaThrGluThrThrAlaGlnAsnG 1071
3203 GAGAA 3207
1071 lyGluValAlaGluGluAlaLysProSerValLysAlaAsnThrGlnThr 1087
3208 ACAGA 3212
1088 AsnGluValAlaGlnSerGlySerGluThrGluGluThrGlnThrThrGl 1104
3213 GGCCGCCTTGACGGCA 3228
1104 uIleLysGluThrAlaLysValGluLysGluGluLysAlaLysValGluL 1121
3228 3228
1121 ysGluGluLysAlaLysValGluLysAspGluIleGlnGluAlaProGln 1137
3229 .....AAACAGGCACAACTTGCCGCCAAA..... 3252
1138 MetAlaSerGluThrSerProLysGlnAlaLysProAlaProLysGluVa 1154
3252 3252
1154 lSerThrAspThrLysValGluGluThrGlnValGlnAlaGlnProGlnT 1171
3252 3252
1171 hrGlnSerThrThrValAlaAlaAlaGluAlaThrSerProAsnSerLys 1187
3253 .....CAACAGGCGGAAAAAGACAACGCGCAAAGCCTTGA 3287
1188 ProAlaGluGluThrGlnProSerGluLysThrAsnAlaGluProValTh 1204
3288 CGCGCTGATTGCGGCCGGGCGCAAT 3312
1204 rProValValSerLysAsnGlnThrGluAsnThrThrAspGlnProThrG 1221
3313 .....GCCACCGAAAAGGCAGAAAGTGTTGCC 3339
1221 luArgGluLysThrAlaLysValGluThrGluLysThrGlnGluProPro 1237
3340 GAACCGGCCCGGCAGGCAGGC 3360
1238 GlnValAlaSerGlnAlaSerProLysGlnGluGlnSerGluThrValGl 1254
3361 .....GGGGAAAATGCC...GGCATTATGCAGGCGG 3388
1254 nProGlnAlaValLeuGluSerGluAsnValProThrValAsnAsnAlaG 1271
3389 AGGAAGAGAAAAAACGGGTGCAGGCGGATAAAGAC 3423
1271 luGluValGlnAlaGlnLeuGlnThrGlnThrSerAlaThrValSerThr 1287


3424 ACCGCCTT 3431
1288 LysGlnProAlaProGluAsnSerIleAsnThrGlySerAlaThrAlaIl 1304
3432 GGCGAAACAGCGCGAAGCGGAAACCCGGCCGGCTACC 3468
1304 eThrGluThrAlaGluLysSerAspLysProGlnThrGluThrAlaAlaS 1321
3468 3468
1321 erThrGluAspAlaSerGlnHisLysAlaAsnThrValAlaAspAsnSer 1337
3469 .....ACCGCCTTCCCCCGCGCCCGCCGCGCCCGCCG 3500
1338 ValAlaAsnAsnSerGluSerSerGluProLysSerArgArg...ArgAr 1353
3501 GGATTTGCCGCAACCGCAG 3519
1353 gSerIleSerGlnProGlnGluThrSerAlaGluGluThrThrAlaAlaS 1370
3520 CCCCAACCGCAA 3531
1370 erThrAspGluThrThrIleAlaAspAsnSerLysArgSerLysProAsn 1386
3532 CCCCAACCGCAG 3543
1387 ArgArgSerArgArgSerValArgSerGluProThrValThrAsnGlySe 1403
3544 .....CGCGACCTGATCAGCCGTTATGCCAATA 3571
1403 rAspArgSerThrValAlaLeuArgAspLeuThrSerThrAsnThrAsnA 1420
3572 GCGGTTTGAGTGAATTTTCCGCC.....ACGCTCAACAGC 3606
1420 laValIleSerAspAlaMetAlaLysAlaGlnPheValAlaLeuAsnVal 1436
3607 GTTTTCGCCGTACAGGACGAATTGGACCGCGTG...TTTGCCGAAGACCG 3653
1437 GlyLysAlaValSerGlnHisIleSerGlnLeuGluMetAsnAsnGluGl 1453
3654 CCGCAACGCCGTTTGGACAAGCGGCATCCGGGACACCAAACACTACCGTT 3703
1453 yGlnTyrAsnValTrpValSerAsnThrSerMetAsnGluAsnTyrSerS 1470
3704 CGCAAGATTTCCGCGCCTACCGCCAACAAACCGACCTGCGCCAAATCGGT 3753
1470 erSerGlnTyrArgArgPheSerSerLysSerThrGlnThrGlnLeuGly 1486
3754 ATGCAGAAAAACCTCGGCAGCGGG...CGCGTCGGCATCCTGTTTTCGCA 3800
1487 TrpAspGlnThrIleSerAsnAsnValGlnLeuGlyGlyValPheThrTy 1503
3801 CAACCGGACCGGAAACACCTTCGACGACGGCATCGGCAACTCGGCACGGC 3850
1503 rValArgAsnSerAsnAsnPheAspLysAlaSerSerLysAsnThrLeuA 1520
3851 TTGCCCACGGTGCCGTTTTCGGGCAATACGGCATCGGCAGGTTCGACATC 3900
1520 laGlnValAsnPheTyrSerLysTyrTyrAlaAspAsnHisTrpTyrLeu 1536
3901 GGCATCAGCGCGGGCGCGGGTTTTAGTAGCGGCAGCCTTTCAGACGGCAT 3950
1537 GlyIleAspLeuGlyTyrGlyLysPheGlnSerAsnLeuLysThrAsnHi 1553
3951 CAGAGGCAAAATCCGCCGCCGCGTGCTGCATTACGGCATTCAGGCAAGAT 4000
1553 sAsnAlaLysPheAlaArgHisThrAlaGlnPheGlyLeuThrAlaGlyL 1570
4001 ACCGCGCAGGTTTCGGCGGATTCGGCATCGAACCGCACATCGGCGCAACG 4050
1570 ysAlaPheAsnLeuGlyAsnPheGlyIleThrProIleValGlyValArg 1586
4051 CGCTATTTCGTCCAAAAAGCGGATTACCGATACGAAAACGTCAATATCGC 4100

```
                ::::::::::::::|||::::::         :::   :::|||
    1587 TyrSerTyrLeuSerAsnAlaAsnPheAlaLeuAlaLysAspArgIleLy 1603

    4101 CACCCCGGGCCTTGCATTCAACCGCTACCGCGCGGGCATTAAGGCAGATT 4150
          :::      ::::::   :::         |||   :::      :::|
    1603 sValAsnProIleSerValLysThrAlaPheAlaGlnValAspLeuSerT 1620

    4151 ATTCATTCAAACCGGCGCAACACATTTCCATCACGCCTTATTTGAGCCTG 4200
         ||::::::      ::::::   :::|||:::||||||   ||||||
    1620 yrThrTyrHisLeuGlyGlu...PheSerValThrProIleLeuSerAla 1635

    4201 TCCTAT...ACCGATGCCGCTTCCGGCAAAGTCCGAACGCGCGTCAATAC 4247
            |||   |||:::   :::||||||||||:::      :::||||||
    1636 ArgTyrAspThrAsnGlnGlySerGlyLysIle......AsnValAsnGl 1650

    4248 CGCCGTATTGGCGCAGGATTTCGGCAAAACCCGCAGTGCGGAATGGGGCG 4297
                 :::|||   :::      :::   ::::::   :::   |||:
    1650 nTyrAspPheAlaTyrAsnValGluAsnGlnGlnGlnTyrAsnAlaGlyL 1667

    4298 TAAACGCCGAAATCAAAGGTTTCACGCTGTCCCTCCACGCTGCCGCCGCC 4347
         :::::   :::      :::      |||||||||   ::::::   :::
    1667 euLysLeuLysTyrHisAsnValLysLeuSerLeuIleGlyGlyLeuThr 1683

    4348 AAGGGGCCGCAATTGGAAGCGCAGCACAGCGCGGGCATCAAATTAGGCTA 4397
         |||:::   |||   |||   |||   :::|||   :::||||||:::::
    1684 LysAlaLysGlnAlaGluLysGlnLysThrAlaGluLeuLysLeuSerPh 1700

    4398 C 4398
         :
    1700 e 1700

seq_name: pir2:A26039

seq_documentation_block:
IgA-specific metalloendopeptidase (EC 3.4.24.13) precursor - Neisseria gonorrhoeae (stra
N;Alternate names: IgA protease; immunoglobulin A1 proteinase
C;Species: Neisseria gonorrhoeae
A;Variety: strain MS11
C;Date: 05-Oct-1988 #sequence_revision 05-Oct-1988 #text_change 08-Dec-2000
C;Accession: A26039; S09386
R;Pohlner, J.; Halter, R.; Beyreuther, K.; Meyer, T.F.
Nature 325, 458-462, 1987
A;Title: Gene structure and extracellular secretion of Neisseria gonorrhoeae IgA proteas
A;Reference number: A26039; MUID:87115823
A;Accession: A26039
A;Molecule type: DNA
A;Residues: 1-1532 <POH>
A;Cross-references: GB:X04835; NID:g44868; PIDN:CAA28538.1; PID:g44869
A;Note: the authors translated the codon AAG for residue 668 as Asn
R;Halter, R.; Pohlner, J.; Meyer, T.F.
EMBO J. 8, 2737-2744, 1989
A;Title: Mosaic-like organization of IgA protease genes in Neisseria gonorrhoeae generat
A;Reference number: S09386; MUID:90060036
A;Accession: S09386
A;Status: not compared with conceptual translation
A;Molecule type: DNA
A;Residues: 281-325,'N',327-337,'N',339-427,'M',429-531,'N',533-615,'V',617-631,'N',633-
A;Experimental source: strain MS11
C;Genetics:
A;Gene: iga
C;Superfamily: IgA-specific metalloendopeptidase
C;Keywords: hydrolase; metalloproteinase; transmembrane protein
F;1-27/Domain: signal sequence #status predicted <SIG>
F;28-1532/Product: immunoglobulin A1 proteinase #status predicted <MAT>
F;986-987/Cleavage site: Pro-Ser (autolytic) #status predicted
F;1018-1019/Cleavage site: Pro-Ser (autolytic) #status predicted
F;1121-1122/Cleavage site: Pro-Ala (autolytic) #status predicted


alignment_scores:
           Quality: 1565.50            Length:   1738
             Ratio:  1.757               Gaps:     46
 Percent Similarity:  51.266   Percent Identity: 27.100

alignment_block:
 US-09-303-518D-653 x A26039   ..

 Align seg 1/1  to: A26039  from: 1  to: 1532

      49 AAAACCGGCCGCATCCGCTTCTCGCCCGCTTACTTAGCCATATGCCTGTC 98
         |||:::   |||:::::::::::::         |||:::|||   |||::
       2 LysAlaLysArgPheLysIleAsnAlaIleSerLeuSerIlePheLeuAl 18

      99 GTTCGGCATTCTGCCCCAAGCCCGGGCGGGACACACTTATTTCGGCATCA 148
         ::::::::::   |||   :::::::|||:::   :::         ::::
      18 aTyrAlaLeuThrProTyrSerGluAlaAlaLeuValArgAspAspValA 35

     149 ACTACCAATACTATCGCGACTTTGCCGAAAATAAAGGCAAGTTTGCAGTC 198
         ::||||||   :::||||||||||||||||||||||||||||||   |||
      35 spTyrGlnIlePheArgAspPheAlaGluAsnLysGlyLysPhePheVal 51

     199 GGGGCGAAAGATATTGAGGTTTACAACAAAAAAGGGGAGTTGGTCGGCAA 248
         ||||||   |||::::::|||   ||||||:::|||:::   :::|||::
      52 GlyAlaThrAspLeuSerValLysAsnLysArgGlyGlnAsnIleGlyAs 68

     249 ATCGATGACGAAAGCCCCGATGATTGATTTTTCTGTGGTATCGCGTAAC. 297
         ::::::::::::::::|||||||||||||||||||||||:::::   |||
      68 nAlaLeuSerAsnValProMetIleAspPheSerValAlaAspValAsnL 85

     298 ..GGCGTGGCGGCATTGGCGGGCGATCAATATATTGTGAGCGTGGCACAT 345
              :::|||:::::::::      ||||||   |||||||||   |||
      85 ysArgIleAlaThrValValAspProGlnTyrAlaValSerValLysHis 101

     346 ..........................AACGGCGGCTATAACAATGTTGA 368
                                     |||   ||||||:::   ::
     102 AlaLysAlaGluValHisThrPheTyrTyrGlyGlnTyrAsnGlyHisAs 118

     369 TTTTGGTGCGGAGGGAAGCAATCCCGATCAGCACCGCTTTTCTTACCAAA 418
         :      |||:::   :::|||                  :::|||::::
     118 nAspValAlaAspLysGluAsn.................GluTyrArgV 129

     419 TTGTGAAAAGAAATAATTATAAAGCAGGGACTAACGGCCATCCTTATGGC 468
         ::|||:::::|||||||||:::                |||
     129 alValGluGlnAsnAsnTyrGluPro............HisLysAlaTrp 141

     469 GGC.......................GATTATCATATGCCGCGTTTGCA 494
         |||                       ||||||:::|||   |||:::::
     142 GlyAlaSerAsnLeuGlyArgLeuGluAspTyrAsnMetAlaArgPheAs 158

     495 CAAATTTGTCACAGATGCAGAACCTGTTGAGATGACCAGTTATATGGATG 544
         :||||||||||||:::::   |||:::      |||:::       |
     158 nLysPheValThrGluValAlaProIleAlaProThrAspAlaGlyGlyG 175

     545 GGTGG...AAATACGCTGATTTAAATAAATACCCTGATCGTGTTCGAATC 591
         ||         |||   |||   |||::::::   :::   |||||||||
     175 lyLeuAspThrTyrLysAspLysAsnArgPheSerSerPheValArgIle 191

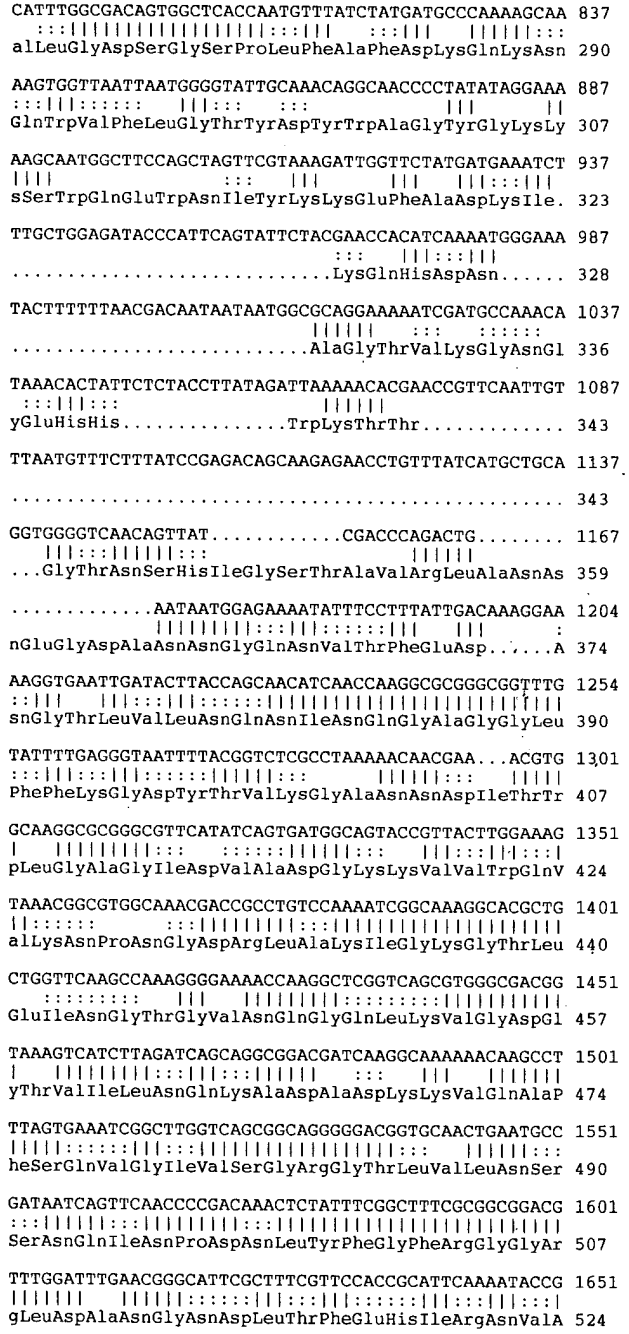
     592 GGAGCAGGCAGACAA......TATTGGCGGTCTGATGAAGACGAACCCAA 635
         ||||||||||||||      |||   ::::::          :::   ::
     192 GlyAlaGlyArgGlnLeuValTyrGluLysGlyValTyrHisGlnGluGl 208

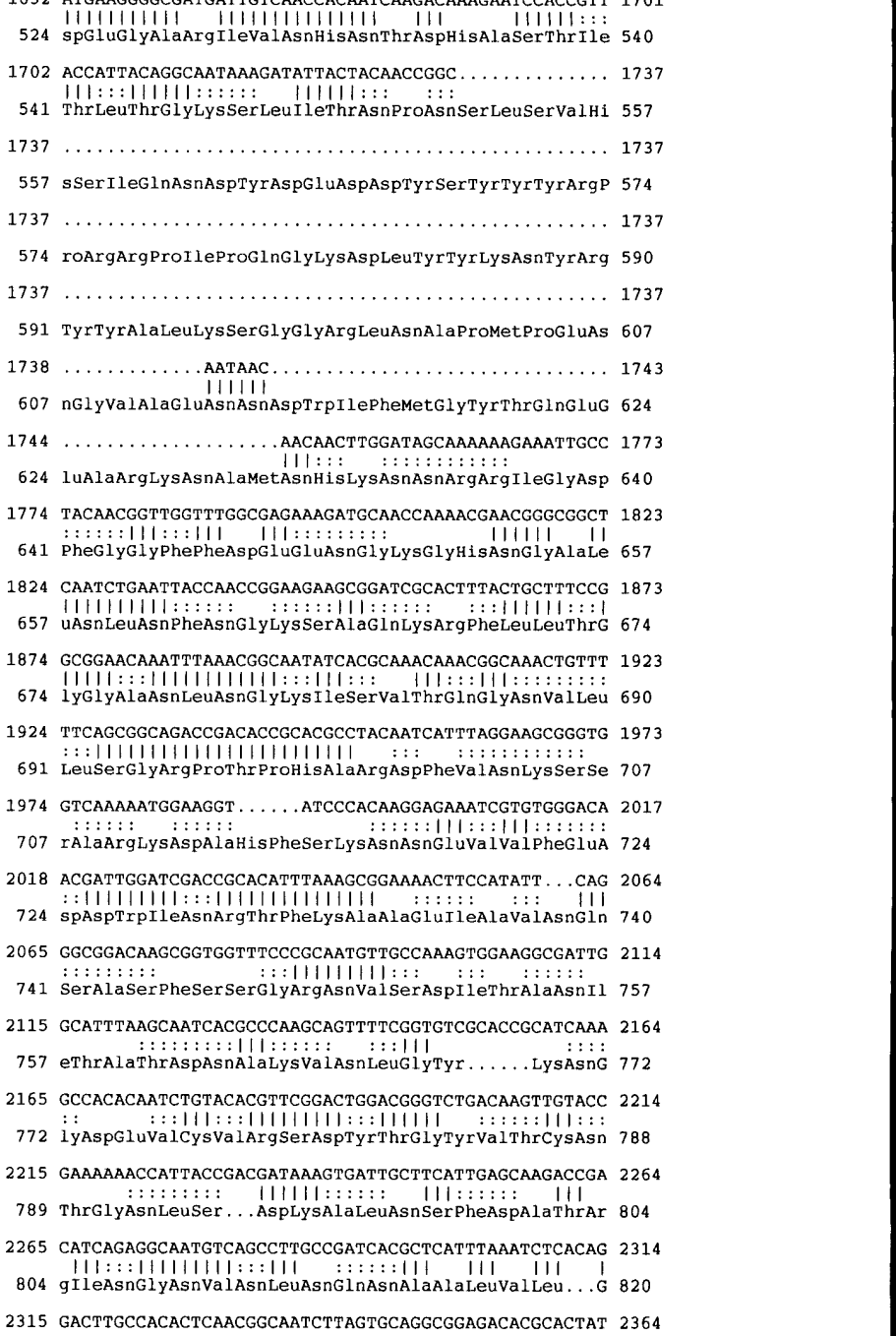
     636 TAACCGCGAAAGTTCATAT.........CATATTGCAAGCGCATATTCTT 676
         :|||   |||::::::|||           :::::::::||||||   :
     208 yAsn...GluLysGlyTyrAspLeuArgAspLeuSerGlnAlaTyrArgT 224

     677 GGCTCGTCGGTGGC................................. 690
         ::   ::::::|||
     224 yrAlaIleAlaGlyThrProTyrLysAspIleAsnIleAspGlnThrMet 240

     691 AATACCTTTGCACAAAATGGATCAGGTGGTGGCACAGTCAACTTAGGTAG 740
         ||||||   :::      |||   |||:::   :::   :::   :::::
     241 AsnThrGluGlyLeuIleGlyPheGlyAsnHisAsnLysGlnTyrSerAl 257

     741 CGAAAAAATTAAACATAGC...CCATATGGTTTTTTACCAACAGGAGGCT 787
         :|||::::::|||::::::               |||   ::   |||
     257 aGluGluLeuLysGlnAlaLeuSerGlnAspAlaLeuThrAsnTyrGlyV 274
```


788 CATTTGGCGACAGTGGCTCACCAATGTTTATCTATGATGCCCAAAAGCAA 837
274 alLeuGlyAspSerGlySerProLeuPheAlaPheAspLysGlnLysAsn 290
838 AAGTGGTTAATTAATGGGGTATTGCAAACAGGCAACCCCTATATAGGAAA 887
291 GlnTrpValPheLeuGlyThrTyrAspTyrTrpAlaGlyTyrGlyLysLy 307
888 AAGCAATGGCTTCCAGCTAGTTCGTAAAGATTGGTTCTATGATGAAATCT 937
307 sSerTrpGlnGluTrpAsnIleTyrLysLysGluPheAlaAspLysIle. 323
938 TTGCTGGAGATACCCATTCAGTATTCTACGAACCACATCAAAATGGGAAA 987
324 .....LysGlnHisAspAsn..... 328
988 TACTTTTTTAACGACAATAATAATGGCGCAGGAAAAATCGATGCCAAACA 1037
329 .....AlaGlyThrValLysGlyAsnGl 336
1038 TAAACACTATTCTCTACCTTATAGATTAAAAACACGAACCGTTCAATTGT 1087
336 yGluHisHis.....TrpLysThrThr..... 343
1088 TTAATGTTTCTTTATCCGAGACAGCAAGAGAACCTGTTTATCATGCTGCA 1137
343 343
1138 GGTGGGGTCAACAGTTAT.....CGACCCAGACTG..... 1167
344 ...GlyThrAsnSerHisIleGlySerThrAlaValArgLeuAlaAsnAs 359
1168 .....AATAATGGAGAAAATATTTCCTTTATTGACAAAGGAA 1204
359 nGluGlyAspAlaAsnAsnGlyGlnAsnValThrPheGluAsp.....A 374
1205 AAGGTGAATTGATACTTACCAGCAACATCAACCAAGGCGCGGGCGGTTTG 1254
374 snGlyThrLeuValLeuAsnGlnAsnIleAsnGlnGlyAlaGlyGlyLeu 390
1255 TATTTTGAGGGTAATTTTACGGTCTCGCCTAAAAACAACGAA...ACGTG 1301
391 PhePheLysGlyAspTyrThrValLysGlyAlaAsnAsnAspIleThrTr 407
1302 GCAAGGCGCGGGCGTTCATATCAGTGATGGCAGTACCGTTACTTGGAAAG 1351
407 pLeuGlyAlaGlyIleAspValAlaAspGlyLysLysValValTrpGlnV 424
1352 TAAACGGCGTGGCAAACGACCGCCTGTCCAAAATCGGCAAAGGCACGCTG 1401
424 alLysAsnProAsnGlyAspArgLeuAlaLysIleGlyLysGlyThrLeu 440
1402 CTGGTTCAAGCCAAAGGGGAAAACCAAGGCTCGGTCAGCGTGGGCGACGG 1451
441 GluIleAsnGlyThrGlyValAsnGlnGlyGlnLeuLysValGlyAspGl 457
1452 TAAAGTCATCTTAGATCAGCAGGCGGACGATCAAGGCAAAAAACAAGCCT 1501
457 yThrValIleLeuAsnGlnLysAlaAspAlaAspLysLysValGlnAlaP 474
1502 TTAGTGAAATCGGCTTGGTCAGCGGCAGGGGGACGGTGCAACTGAATGCC 1551
474 heSerGlnValGlyIleValSerGlyArgGlyThrLeuValLeuAsnSer 490
1552 GATAATCAGTTCAACCCCGACAAACTCTATTTCGGCTTTCGCGGCGGACG 1601
491 SerAsnGlnIleAsnProAspAsnLeuTyrPheGlyPheArgGlyGlyAr 507
1602 TTTGGATTTGAACGGGCATTCGCTTTCGTTCCACCGCATTCAAAATACCG 1651
507 gLeuAspAlaAsnGlyAsnAspLeuThrPheGluHisIleArgAsnValA 524


1652 ATGAAGGGGCGATGATTGTCAACCACAATCAAGACAAAGAATCCACCGTT 1701
524 spGluGlyAlaArgIleValAsnHisAsnThrAspHisAlaSerThrIle 540
1702 ACCATTACAGGCAATAAAGATATTACTACAACCGGC..... 1737
541 ThrLeuThrGlyLysSerLeuIleThrAsnProAsnSerLeuSerValHi 557
1737 1737
557 sSerIleGlnAsnAspTyrAspGluAspAspTyrSerTyrTyrTyrArgP 574
1737 1737
574 roArgArgProIleProGlnGlyLysAspLeuTyrTyrLysAsnTyrArg 590
1737 1737
591 TyrTyrAlaLeuLysSerGlyGlyArgLeuAsnAlaProMetProGluAs 607
1738 .....AATAAC..... 1743
607 nGlyValAlaGluAsnAsnAspTrpIlePheMetGlyTyrThrGlnGluG 624
1744 .....AACAACTTGGATAGCAAAAAAGAAATTGCC 1773
624 luAlaArgLysAsnAlaMetAsnHisLysAsnAsnArgArgIleGlyAsp 640
1774 TACAACGGTTGGTTTGGCGAGAAAGATGCAACCAAAACGAACGGGCGGCT 1823
641 PheGlyGlyPhePheAspGluGluAsnGlyLysGlyHisAsnGlyAlaLe 657
1824 CAATCTGAATTACCAACCGGAAGAAGCGGATCGCACTTTACTGCTTTCCG 1873
657 uAsnLeuAsnPheAsnGlyLysSerAlaGlnLysArgPheLeuLeuThrG 674
1874 GCGGAACAAATTTAAACGGCAATATCACGCAAACAAACGGCAAACTGTTT 1923
674 lyGlyAlaAsnLeuAsnGlyLysIleSerValThrGlnGlyAsnValLeu 690
1924 TTCAGCGGCAGACCGACACCGCACGCCTACAATCATTTAGGAAGCGGGTG 1973
691 LeuSerGlyArgProThrProHisAlaArgAspPheValAsnLysSerSe 707
1974 GTCAAAAATGGAAGGT......ATCCCACAAGGAGAAATCGTGTGGGACA 2017
707 rAlaArgLysAspAlaHisPheSerLysAsnAsnGluValValPheGluA 724
2018 ACGATTGGATCGACCGCACATTTAAAGCGGAAAACTTCCATATT...CAG 2064
724 spAspTrpIleAsnArgThrPheLysAlaAlaGluIleAlaValAsnGln 740
2065 GGCGGACAAGCGGTGGTTTCCCGCAATGTTGCCAAAGTGGAAGGCGATTG 2114
741 SerAlaSerPheSerSerGlyArgAsnValSerAspIleThrAlaAsnIl 757
2115 GCATTTAAGCAATCACGCCCAAGCAGTTTTCGGTGTCGCACCGCATCAAA 2164
757 eThrAlaThrAspAsnAlaLysValAsnLeuGlyTyr......LysAsnG 772
2165 GCCACACAATCTGTACACGTTCGGACTGGACGGGTCTGACAAGTTGTACC 2214
772 lyAspGluValCysValArgSerAspTyrThrGlyTyrValThrCysAsn 788
2215 GAAAAAACCATTACCGACGATAAAGTGATTGCTTCATTGAGCAAGACCGA 2264
789 ThrGlyAsnLeuSer...AspLysAlaLeuAsnSerPheAspAlaThrAr 804
2265 CATCAGAGGCAATGTCAGCCTTGCCGATCACGCTCATTTAAATCTCACAG 2314
804 gIleAsnGlyAsnValAsnLeuAsnGlnAsnAlaAlaLeuValLeu...G 820
2315 GACTTGCCACACTCAACGGCAATCTTAGTGCAGGCGGAGACACGCACTAT 2364

```
     ||   |||:::|||   |||:::::::::::::
 820 lyLysAlaAlaLeuTrpGlyLysIleGlnGly..... 830

2365 ACGGTTACGCGCAACGCCACCCAAAACGGCAACCTCAGCCTCGTGGGCAA 2414
                                         |||||
 831 .....GlnGlyAs 833

2415 TGCCCAAGCAACATTTAATCAAGCCACATTAAACGGCAACACATCGGCTT 2464
     |:::::::::::::::|||||
 833 nSerArgValSerLeuAsnGln..... 840

2465 CGGACAATGCTTCATTTAATCTAAGCAACAACGCCGTACAAAACGGCAGT 2514

 840 ..... 840

2515 CTGACGCTTTCCGACAACGCTAAGGCAAACGTAAGCCATTCCGCACTCAA 2564

 840 ..... 840

2565 CGGCAATGTCTCCCTAGCCGATAAGGCAGTATTCCATTTTGAAAACAGCC 2614

 840 ..... 840

2615 GCTTTACCGGAAAAATCAGCGGCGGCAAGGATACGGCATTACACTTAAAA 2664

 840 ..... 840

2665 GACAGCGAATGGACGCTGCCGTCGGGCACGGAATTAGGCAATTTAAACCT 2714
        |||:::|||   |||   :::   :::::::::   ||||||:::||
 841 HisSerLysTrpHisLeuThrGlyAspSerGlnValHisAsnLeuSerLe 857

2715 TGACAACGCCACCATTACACTCAATTCCGCCTATCGACACGATGCGGCAG 2764
     |   ::::::   |||   ||||||:::|||         ||||||   :
 857 uAlaAspSerHisIleHisLeuAsnAsnAla......SerAspAlaGlnS 872

2765 GCGCGCAAACCGGCAGTGCGGCAGATGCGCCGCGCCGCCGTTCGCGCCGT 2814
     ::|||
 872 erAla..... 873

2815 TCCCTATTATCCGTTACGCCGCCAACTTCGGCAGAATCCCGTTTCAACAC 2864
                                        ::::::::::::||
 874 .....AsnLysTyrHisTh 878

2865 GCTGACGGTAAACGGCAAATTGAACGGTCAGGGAACATTCCGCTTTATGT 2914
     |:::   :::|||      |||:::|||:::|||   |||::::::::::
 878 rIleLysIleAsn...HisLeuSerGlyAsnGlyHisPheHisTyrLeuT 894

2915 CGGAACTCTTCGGCTACCGCAGCGGCAAATTGAAGCTGGCGGAAAGTTCC 2964
     :::::|||           :::   |||:::   :::   ||||||:::
 894 hrAspLeuAlaLysAsnLeuGlyAspLysValLeuValLysGluSerAla 910

2965 GAAGGCACTTACACCTTGGCTGTCAACAATACCGGCAACGAACCCGTAAG 3014
     :::|||   |||   |||   |||:::|||      :::||||||   ::
 911 SerGlyHisTyrGlnLeuHisValGlnAsnLysThrGlyGluProAsnGl 927

3015 TCTCGAGCAATTGACGGTAGTGGAAGGAAAAGACAACACACCGCTGTCC. 3063
     :                    ||||||   |||           |||
 927 n.....GluGlyLeuAspLeuPheAspAlaSerS 937

3064 .....GAAAATCTTAATTTCACCCTGCAAAACGAACACGTC 3099
                 ::::::|||      :::|||   |||::::::|||
 937 erValGlnAspArgSerArgLeuPheValSerLeuAlaAsnHisTyrVal 953

3100 GATGCCGGCGCATGGCGTTATCAGCTTATCCGCAAAGACGGCGAGTTCCG 3149
     |||   ||||||   ||||||   :::      ::::::|||      ||
 954 AspLeuGlyAlaLeuArgTyrThrIleLysThrGluAsnGlyIleThrAr 970

3150 CCTGCATAAT.....CCGGTCAAA..... 3168
     ||||:::|||                |||||||||
 970 gLeuTyrAsnProTyrAlaGlyAsnGlyArgProValLysProAlaProS 987

3169 .....GAACAAGAGCTTTCCGACAAACTCGGCAAGGCGGGAGAAACAGAG 3213
               :::   :::::::::      ||||||   :::|||:::
```

```
 987 erProAlaAlaAsnThrAlaSerGlnAlaGlnLysAlaThrGlnThrAsp 1003

3214 GCCGCCTTGACGGCAAAA..... 3231
     :::|||      ||||||
1004 GlyAlaGlnIleAlaLysProGlnAsnIleValValAlaProProSerPr 1020

3232 .....CAGGCACAACTTGCCGCCAAACAACAGGCGGAAAAAGACA 3271
               ||||||:::   |||   :::||||||||||:::   ::::
1020 oGlnAlaAsnGlnAlaGluGluAlaLeuArgGlnGlnAlaLysAlaGluG 1037

3272 ACGCG.....CAAAGCCTTGACGCGCTGATTGCGGCCGGGCGCAAT 3312
     :::::         |||:::   :::|||      :::|||   ::::::
1037 lnValLysArgGlnGlnAlaAlaGluAlaGluLysValAlaArgGlnLys 1053

3313 GCCACCGAAAAGGCAGAAAGTGTTGCCGAACCGGCCCGGCAG..... 3354
           |||      :::::::::||||||   |||||||||
1054 AspGluGluAlaLysArgLysAlaAlaGluIleAlaArgGlnGlnGluGl 1070

3355 .......GCAGGCGGGGAAAATGCCGGCATTATGCAGGCGGAGGAAGAGA 3397
               :::::::|||   |||:::   ::::::||||||   |||:
1070 uAlaArgLysAlaAlaGluLeuAlaAlaLysGlnLysAlaGluAlaGluA 1087

3398 AAAAACGGGTGCAGGCGGATAAAGACACCGCCTTGGCGAAACAGCGCGAA 3447
     ::|||      ::                    ||||||:::|||:::
1087 rgLysAlaArgGlu.....LeuAlaArgGlnLysAla 1097

3448 GCGGAAACCCGGCCGGCTACCACCGCCTTCCCCCGCGCCCGCCGCGCCCG 3497
        |||:::         ::::::|||   |||:::   ||||||   ||
1098 GluGluAlaSerHisGlnAlaAsnAlaLysProLysArgArgArgArgAr 1114

3498 CCGGGGATTTGCCGCAACCG..... 3516
     |      ||||||:::|||
1114 gAlaIleLeuProArgProProAlaProValPheSerLeuAspAspTyrA 1131

3516 ..... 3516

1131 spAlaLysAspAsnSerGluSerSerIleGlyAsnLeuAlaArgValIle 1147

3516 ..... 3516

1148 ProArgMetGlyArgGluLeuIleAsnAspTyrGluGluIleProLeuGl 1164

3516 ..... 3516

1164 uGluLeuGluAspGluAlaGluGluGluArgArgGlnAlaThrGlnPheH 1181

3516 ..... 3516

1181 isSerLysSerArgAsnArgArgAlaIleSerSerGluProSerSerAsp 1197

3516 ..... 3516

1198 GluAspAlaSerGluSerValSerThrSerAspLysHisProGlnAspAs 1214

3516 ..... 3516

1214 nThrGluLeuHisGluLysValGluThrAlaGlyLeuGlnProArgAlaA 1231

3517 ..CAGCCCCAACCGCAACCCCAACCGCAGCGCGACCTGATCAGCCGTTAT 3564
       ||||||:::   |||         |||   |||   :::|||
1231 laGlnProArgThrGlnAlaAlaAlaGlnAlaAspAlaValSerThrAsn 1247

3565 GCCAATAGCGGTTTGAGTGAATTTTCCGCCACGCTCAACAGCGTT..... 3609
     :::||||||:::||||||:::      |||:::   :::|||:::
1248 ThrAsnSerAlaLeuSerAspAlaMetAlaSerThrGlnSerIleLeuLe 1264

3610 .....TTCGCCGTACAGGACGAATTGGACCGCGTGTTTGCCG 3646
                  :::   ::::::         ::::::      |||:
1264 uAspThrGlyAlaTyrLeuThrArgHisIleAlaGlnLysSerArgAlaA 1281

3647 AAGACCGCCGCAACGCCGTTTGGACAAGCGGCATCCGGGACACCAAACAC 3696
     ::   ::::::|||:::|||||||   |||:::         :::
1281 spAlaGluLysAsnSerValTrpMetSerAsnThrGlyTyrGlyArgAsp 1297
```

```
 3697 TACCGTTCGCAAGATTTCCGCGCCTACCGCCAACAAACCGACCTGCGCCA 3746
      |||   |||   :::::::|||   :::   ::::::           ||
 1298 TyrAlaSerAlaGlnTyrArgArgPheSerSerLysArgThrGlnThrGl 1314

 3747 AATCGGTATGCAGAAAAACCTCGGCAGCGGG...CGCGTCGGCATCCTGT 3793
      |||||||:::::::::::::|||:::::::::   ::::::|||   ::::
 1314 nIleGlyIleAspArgSerLeuSerGluAsnMetGlnIleGlyGlyValL 1331

 3794 TTTCGCACAACCGGACCGGAAACACCTTCGACGACGGCATCGGCAACTCG 3843
      ::::::::::   :::   :::||||||||||::::::   |||::::::
 1331 euThrTyrSerAspSerGlnHisThrPheAspGlnAla...GlyGlyLys 1346

 3844 GCACGGCTTGCCCACGGTGCCGTTTTCGGGCAATACGGCATCGGC...AG 3890
           ::::::::::::   :::::::|||:::|||   ::::::
 1347 AsnThrPheValGlnAlaAsnLeuTyrGlyLysTyrTyrLeuAsnAspAl 1363

 3891 GTTCGACATCGGCATCAGCGCGGGCGCGGGTTTTAGTAGCGGCAGCCTTT 3940
       :::   :::::::   :::   |||||||||          :::   |||:
 1363 aTrpTyrValAlaGlyAspIleGlyAlaGlySerLeuArgSerArgLeuG 1380

 3941 CAGACGGCATCAGAGGCAAAATCCGCCGCCGCGTGCTGCATTACGGCATT 3990
      ::          ::::::::::::::::|||      ::::::   |||:::
 1380 lnThrGlnGlnLysAlaAsnPheAsnArgThrSerIleGlnThrGlyLeu 1396

 3991 CAGGCAAGATACCGCGCAGGTTTCGGCGGATTCGGCATCGAACCGCACAT 4040
                          ::::::   |||   |||   |||
 1397 ThrLeuGlyAsnThrLeuLysIleAsnGlnPheGluIleValProSerAl 1413

 4041 CGGCGCAACGCGCTATTTCGTCCAAAAAGCGGATTACCGATACGAAAACG 4090
       |||                :::::::::|||||||||:::      :::
 1413 aGlyIleArgTyrSerArgLeuSerSerAlaAspTyrLysLeuGlyAspA 1430

 4091 TCAATATCGCCACCCCGGGCCTTGCATTCAACCGCTACCGCGCGGGCATT 4140
        ::::::   :::   ::::::|||   :::          ||||||:::
 1430 spSerValLysValSerSerMetAlaValLysThrLeuThrAlaGlyLeu 1446

 4141 AAGGCAGATTATTCATTCAAACCGGCGCAACACATTTCCATCACGCCTTA 4190
             |||   ||||||   :::   ::::::::::::   |||
 1447 AspPheAlaTyrArgPheLys...ValGlyAsnLeuThrValLysProLe 1462

 4191 TTTGAGCCTGTCCTATACCGATGCCGCTTCCGGCAAAGTCCGAACGCGCG 4240
       ||||||   :::|||      |||      ||||||      :::::::|
 1462 uLeuSerAlaAlaTyrPhe...AlaAsnTyrGlyLysGlyGlyValAsnV 1478

 4241 TCAATACCGCCGTATTGGCGCAGGATTTCGGCAAAACCCGCAGTGCGGAA 4290
      ||:::         :::|||          :::   ::::::   :::
 1478 alGlyGlyLysSerPheAlaTyrLysAlaAspAsnGlnGlnGlnTyrSer 1494

 4291 TGGGGCGTAAACGCCGAAATCAAAGGTTTCACGCTGTCCCTCCACGCTGC 4340
        ||||||          ::::::   ||||||:::::::::::::
 1495 AlaGlyValAlaLeuLeuTyrArgAsnValThrLeuAsnValAsnGlySe 1511

 4341 CGCCGCCAAGGGGCCGCAATTGGAAGCGCAGCACAGCGCGGGCATCAAAT 4390
      :  :::||||||   |||||||||   |||   |||:::   ||||||:
 1511 rIleThrLysGlyLysGlnLeuGluLysGlnLysSerGlyGlnIleLysI 1528

 4391 TAGGCTACCGCTGG 4404
      ::      |||:::
 1528 leGlnIleArgPhe 1532

seq_name: pir2:C41859

seq_documentation_block:
IgA-specific metalloendopeptidase (EC 3.4.24.13) type 1 precursor - Haemophilus influenz
C;Species: Haemophilus influenzae
A;Variety: strain HK613
C;Date: 04-Mar-1993 #sequence_revision 18-Nov-1994 #text_change 08-Dec-2000
C;Accession: C41859
R;Poulsen, K.; Reinholdt, J.; Kilian, M.
J. Bacteriol. 174, 2913-2921, 1992
A;Title: A comparative genetic study of serologically distinct Haemophilus influenzae ty
A;Reference number: A41859; MUID:92234949
A;Accession: C41859
A;Status: preliminary; not compared with conceptual translation
A;Molecule type: nucleic acid
A;Residues: 1-1849 <POU>
A;Experimental source: strain HK613
A;Note: sequence extracted from NCBI backbone (NCBIP:97285)
C;Superfamily: IgA-specific metalloendopeptidase
C;Keywords: hydrolase; metalloproteinase


alignment_scores:
            Quality: 1515.50              Length:   1917
              Ratio:  1.522                Gaps:     60
 Percent Similarity:  51.956   Percent Identity: 25.456

alignment_block:
 US-09-303-518D-653 x C41859   ..

 Align seg 1/1  to: C41859  from: 1  to: 1849

   64 CGCTTCTCGCCCGCTTACTTAGCCATATGCCTGTCGTTCGGCATTCTGCC 113
      :::|||:::      :::::::|||:::   ::::::::::::::   ||
    5 LysPheLysLeuAsnPheIleAlaLeuThrValAlaTyrAlaLeuThrPr 21

  114 CCAAGCCCGGGCGGGACACACTTATTTCGGCATCAACTACCAATACTATC 163
      |   ::::::|||:::   :::          ::::::||||||   :::|
   21 oTyrThrGluAlaAlaLeuValArgAspAspValAspTyrGlnIlePheA 38

  164 GCGACTTTGCCGAAAATAAAGGCAAGTTTGCAGTCGGGGCGAAAGATATT 213
      ||||||||||||||||||||||||||||:::||||||||||   ::::::
   38 rgAspPheAlaGluAsnLysGlyLysPheSerValGlyAlaThrAsnVal 54

  214 GAGGTTTACAACAAAAAAGGGGAGTTGGTCGGCAAATCGATGACGAAAGC 263
      ||||||   :::|||||::::::   :::|||:::::::::   :::::
   55 GluValArgAspLysLysAsnGlnSerLeuGlySerAlaLeuProAsnGl 71

  264 C...CCGATGATTGATTTTTCTGTGGTATCG...CGTAACGGCGTGGCGG 307
      :   |||||||||||||||||||||||||:::      :::   :::|||:
   71 yIleProMetIleAspPheSerValValAspValAspLysArgIleAlaT 88

  308 CATTGGCGGGCGATCAATATATTGTGAGCGTGGCACAT...AACGGCGGC 354
      ::|||::::::   ||||||:::|||:::|||   |||   :::::|||
   88 hrLeuValAsnProGlnTyrValValGlyValLysHisValSerAsnGly 104

  355 TATAACAATGTTGATTTTGGT.............GCGGAGGGAAGCAA 389
        :::::::::   ||||||                :::::::::||
  105 ValSerGluLeuHisPheGlyAsnLeuAsnGlyAsnMetAsnAsnGlyAs 121

  390 TCCCGATCAGCACCGC.................TTTTCTTACCAAATTG 421
      |      :::||||||                      |||      |
  121 nAlaLysSerHisArgAspValSerSerGluGluAsnArgTyrTyrThrV 138

  422 TGAAAAGAAATAATTAT.............................AAA 441
      ||::::::||||||:::                             |||
  138 alGluLysAsnAsnProThrGluAsnValThrSerPheThrThrLys 154

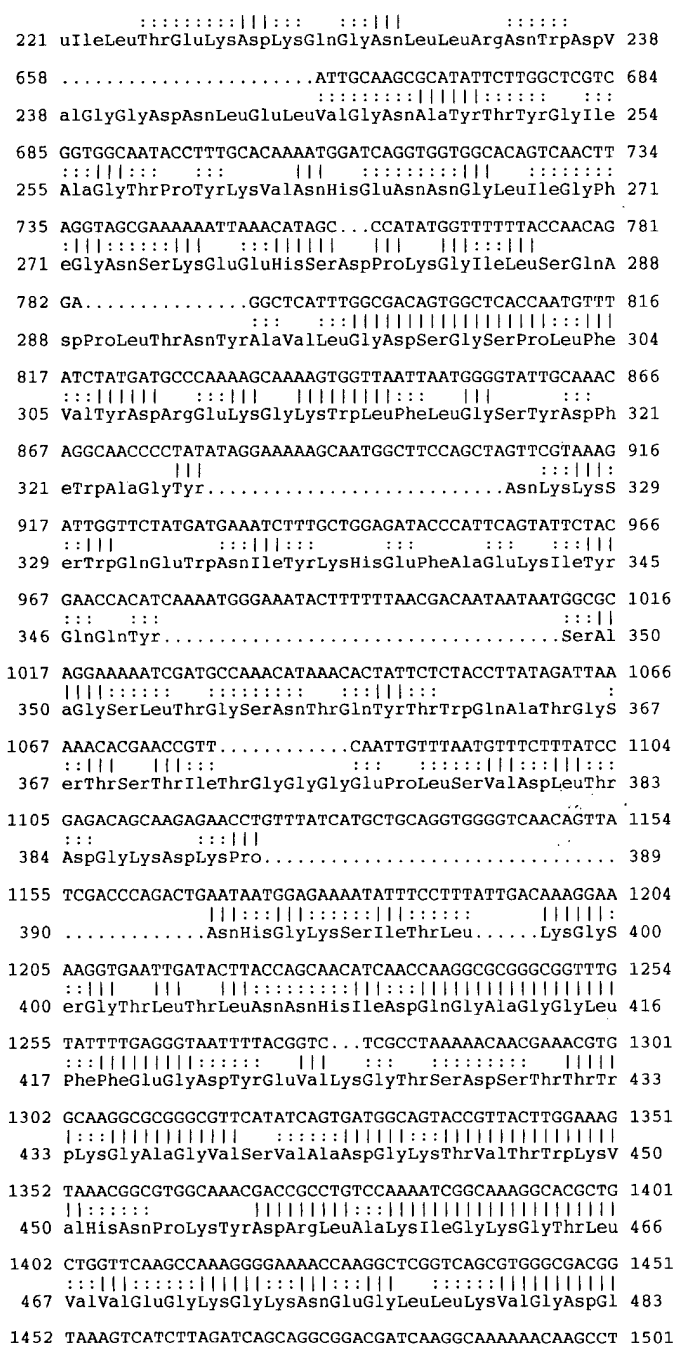
  442 GCAGGGACTAACGGCCATCCTTATGGCGGCGATTATCATATGCCGCGTTT 491
            :::::::::              ||||||:::||||||||||
  155 GluGluGlnAspAlaGlnLysArgArgGluAspTyrTyrMetProArgLe 171

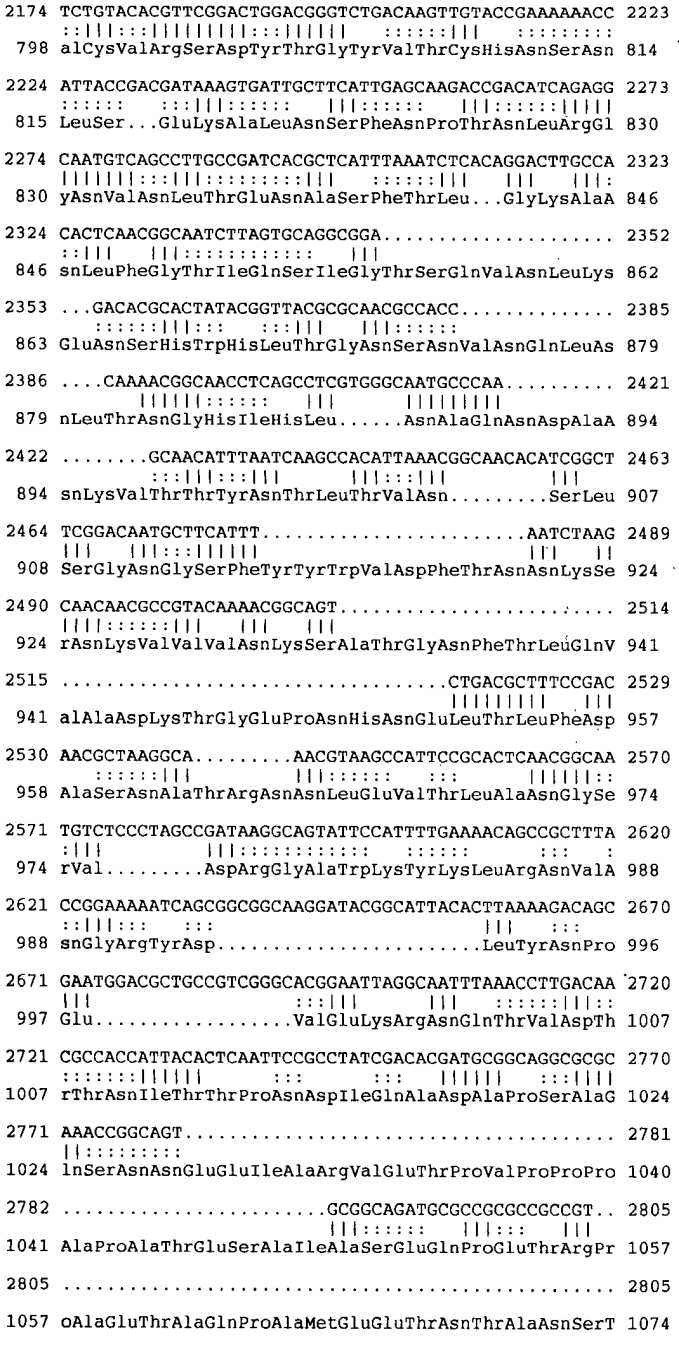
  492 GCACAAATTTGTCACAGATGCAGAACCTGTTGAG...ATGACCAGTTATA 538
      |   ||||||||||:::::   |||:::|||      |||:::
  171 uAspLysPheValThrGluValAlaProIleGluAlaSerThrAlaAsnA 188

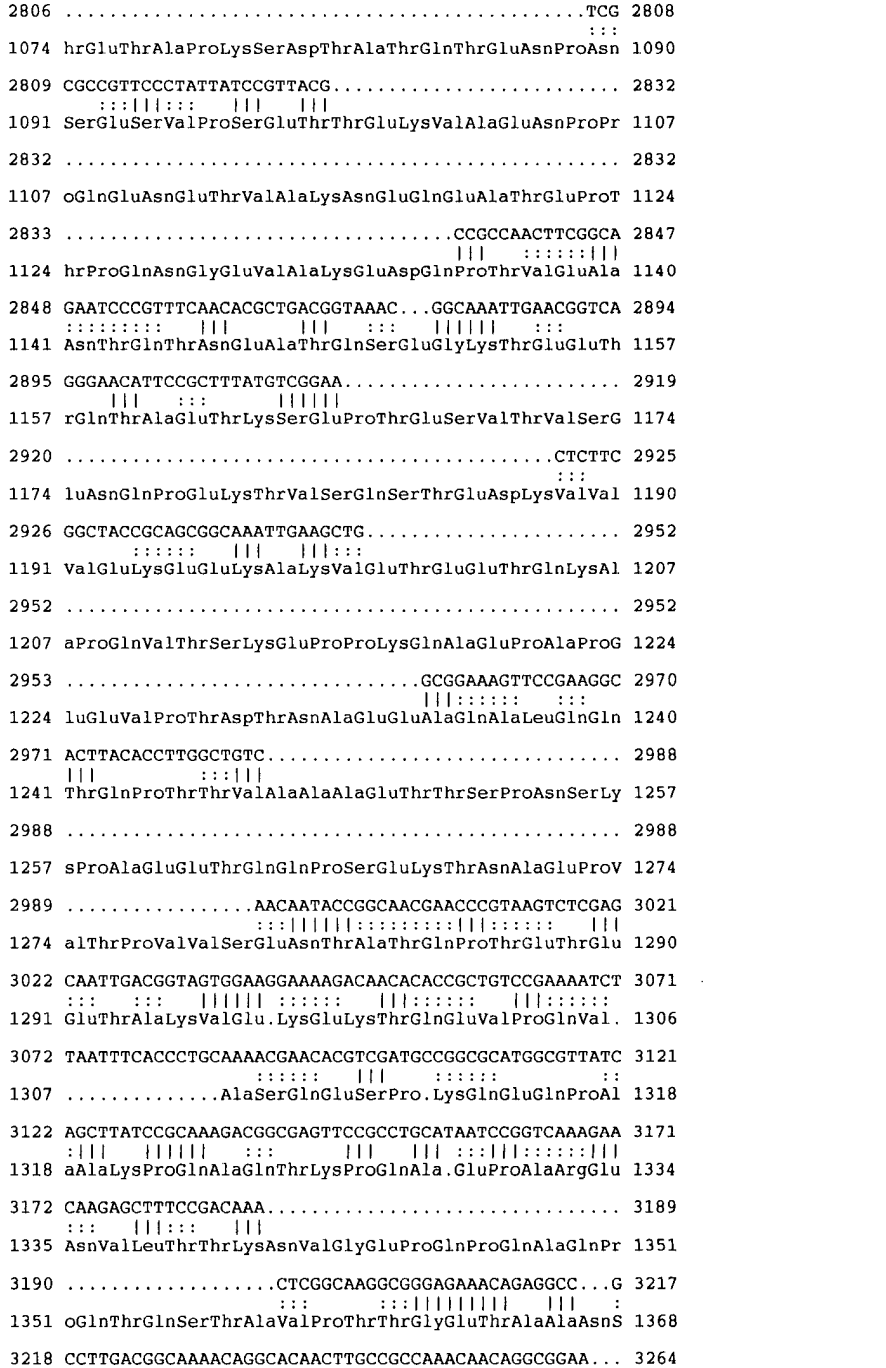
  539 TGGATGGGTGGAAATACGCTGATTTAAATAAATACCCTGATCGTGTTCGA 588
        :::      :::|||   :::   :::|||||||||      ||||||
  188 snAsnLysGlyGluTyrAsnAsnSerAspLysTyrProAlaPheValArg 204

  589 ATCGGAGCAGGCAGACAATAT........................... 609
      :::|||:::|||   |||:::
  205 LeuGlySerGlySerGlnPheIleTyrLysLysGlySerArgTyrGlnLe 221

  610 .TGGCGGTCTGATGAAGACGAACCCAATAACCGCGAAAGTTCATATCAT. 657
```


221 uIleLeuThrGluLysAspLysGlnGlyAsnLeuLeuArgAsnTrpAspV 238
658 .....ATTGCAAGCGCATATTCTTGGCTCGTC 684
238 alGlyGlyAspAsnLeuGluLeuValGlyAsnAlaTyrThrTyrGlyIle 254
685 GGTGGCAATACCTTTGCACAAAATGGATCAGGTGGTGGCACAGTCAACTT 734
255 AlaGlyThrProTyrLysValAsnHisGluAsnAsnGlyLeuIleGlyPh 271
735 AGGTAGCGAAAAAATTAAACATAGC...CCATATGGTTTTTTACCAACAG 781
271 eGlyAsnSerLysGluGluHisSerAspProLysGlyIleLeuSerGlnA 288
782 GA.....GGCTCATTTGGCGACAGTGGCTCACCAATGTTT 816
288 spProLeuThrAsnTyrAlaValLeuGlyAspSerGlySerProLeuPhe 304
817 ATCTATGATGCCCAAAAGCAAAAGTGGTTAATTAATGGGGTATTGCAAAC 866
305 ValTyrAspArgGluLysGlyLysTrpLeuPheLeuGlySerTyrAspPh 321
867 AGGCAACCCCTATATAGGAAAAAGCAATGGCTTCCAGCTAGTTCGTAAAG 916
321 eTrpAlaGlyTyr.....AsnLysLysS 329
917 ATTGGTTCTATGATGAAATCTTTGCTGGAGATACCCATTCAGTATTCTAC 966
329 erTrpGlnGluTrpAsnIleTyrLysHisGluPheAlaGluLysIleTyr 345
967 GAACCACATCAAAATGGGAAATACTTTTTTAACGACAATAATAATGGCGC 1016
346 GlnGlnTyr.....SerAl 350
1017 AGGAAAAATCGATGCCAAACATAAACACTATTCTCTACCTTATAGATTAA 1066
350 aGlySerLeuThrGlySerAsnThrGlnTyrThrTrpGlnAlaThrGlyS 367
1067 AAACACGAACCGTT.....CAATTGTTTAATGTTTCTTTATCC 1104
367 erThrSerThrIleThrGlyGlyGlyGluProLeuSerValAspLeuThr 383
1105 GAGACAGCAAGAGAACCTGTTTATCATGCTGCAGGTGGGGTCAACAGTTA 1154
384 AspGlyLysAspLysPro..... 389
1155 TCGACCCAGACTGAATAATGGAGAAAATATTTCCTTTATTGACAAAGGAA 1204
390 .....AsnHisGlyLysSerIleThrLeu.....LysGlyS 400
1205 AAGGTGAATTGATACTTACCAGCAACATCAACCAAGGCGCGGGCGGTTTG 1254
400 erGlyThrLeuThrLeuAsnAsnHisIleAspGlnGlyAlaGlyGlyLeu 416
1255 TATTTTGAGGGTAATTTTACGGTC...TCGCCTAAAAACAACGAAACGTG 1301
417 PhePheGluGlyAspTyrGluValLysGlyThrSerAspSerThrThrTr 433
1302 GCAAGGCGCGGGCGTTCATATCAGTGATGGCAGTACCGTTACTTGGAAAG 1351
433 pLysGlyAlaGlyValSerValAlaAspGlyLysThrValThrTrpLysV 450
1352 TAAACGGCGTGGCAAACGACCGCCTGTCCAAAATCGGCAAAGGCACGCTG 1401
450 alHisAsnProLysTyrAspArgLeuAlaLysIleGlyLysGlyThrLeu 466
1402 CTGGTTCAAGCCAAAGGGGAAAACCAAGGCTCGGTCAGCGTGGGCGACGG 1451
467 ValValGluGlyLysGlyLysAsnGluGlyLeuLeuLysValGlyAspGl 483
1452 TAAAGTCATCTTAGATCAGCAGGCGGACGATCAAGGCAAAAAACAAGCCT 1501
483 yThrValIleLeuLysGlnLysAlaAspAlaAsnAsnLysValGlnAlaP 500
1502 TTAGTGAAATCGGCTTGGTCAGCGGCAGGGGGACGGTGCAACTGAATGCC 1551
500 heSerGlnValGlyIleValSerGlyArgSerThrLeuValLeuAsnAsp 516
1552 GATAATCAGTTCAACCCCGACAAACTCTATTTCGGCTTTCGCGGCGGACG 1601
517 AspLysGlnValAspProAsnSerIleTyrPheGlyPheArgGlyGlyAr 533
1602 TTTGGATTTGAACGGGCATTCGCTTTCGTTCCACCGCATTCAAAATACCG 1651
533 gLeuAspLeuAsnGlyAsnSerLeuThrPheAspHisIleArgAsnIleA 550
1652 ATGAAGGGGCGATGATTGTCAACCACAATCAAGACAAAGAATCCACCGTT 1701
550 spAspGlyAlaArgValValAsnHisAsnMetThrAsnThrSerAsnIle 566
1702 ACCATTACAGGCAATAAAGATATTACTACA.....ACCGGCAA 1739
567 ThrIleThrGlyGluSerLeuIleThrAsnProAsnThrIleThrSerTy 583
1740 TAAC..... 1743
583 rAsnIleGluAlaGlnAspAspAspHisProLeuArgIleArgSerIleP 600
1743 1743
600 roTyrArgGlnLeuTyrPheAsnGlnAspAsnArgSerTyrTyrThrLeu 616
1743 1743
617 LysLysGlyAlaSerThrArgSerGluLeuProGlnAsnSerGlyGluSe 633
1743 1743
633 rAsnGluAsnTrpLeuTyrMetGlyArgThrSerAspGluAlaLysArgA 650
1744 .....AACAACTTGGATAGCAAAAAAGAAATTGCCTACAACGGTTGG 1785
650 snValMetAsnHisIleAsnAsnGluArgMetAsnGlyPheAsnGlyTyr 666
1786 TTTGGCGAGAAAGATGCAACCAAAACG...AACGGGCGGCTCAATCTGAA 1832
667 PheGlyGluGluGluThrLysAlaThrGlnAsnGlyLysLeuAsnValTh 683
1833 TTACCAACCGGAAGAAGCGGATCGCACTTTACTGCTTTCCGGCGGAACAA 1882
683 rPheAsnGlyLysSerAspGlnAsnArgPheLeuLeuThrGlyGlyThrA 700
1883 ATTTAAACGGCAATATCACGCAAACAAACGGCAAACTGTTTTTCAGCGGC 1932
700 snLeuAsnGlyAspLeuAsnValGluLysGlyThrLeuPheLeuSerGly 716
1933 AGACCGACACCGCACGCCTACAATCATTTAGGAAGCGGGTGGTCAAAAAT 1982
717 ArgProThrProHisAlaArgAspIleAlaGlyIleSerSerThrLysLy 733
1983 GGAAGGT.....ATCCCACAAGGAGAAATCGTGTGGGACAACGATTGGA 2026
733 sAspProHisPheThrGluAsnAsnGluValValValGluAspAspTrpI 750
2027 TCGACCGCACATTTAAAGCGGAAAACTTCCATATTCAGGGCGGACAAGCG 2076
750 leAsnArgAsnPheLysAlaThrThrMetAsnValThrGlyAsnAlaSer 766
2077 GTGGTTTCC...CGCAATGTTGCCAAAGTGGAAGGCGATTGGCATTTAAG 2123
767 LeuTyrSerGlyArgAsnValAlaAsnIleThrSerAsnIleThrAlaSe 783
2124 CAATCACGCCCAAGCAGTTTTCGGTGTCGCACCGCATCAAAGCCACACAA 2173
783 rAsnAsnAlaGlnValHisIleGlyTyr.....LysThrGlyAspThrV 798


2174 TCTGTACACGTTCGGACTGGACGGGTCTGACAAGTTGTACCGAAAAAACC 2223
798 alCysValArgSerAspTyrThrGlyTyrValThrCysHisAsnSerAsn 814
2224 ATTACCGACGATAAAGTGATTGCTTCATTGAGCAAGACCGACATCAGAGG 2273
815 LeuSer...GluLysAlaLeuAsnSerPheAsnProThrAsnLeuArgGl 830
2274 CAATGTCAGCCTTGCCGATCACGCTCATTTAAATCTCACAGGACTTGCCA 2323
830 yAsnValAsnLeuThrGluAsnAlaSerPheThrLeu...GlyLysAlaA 846
2324 CACTCAACGGCAATCTTAGTGCAGGCGGA..... 2352
846 snLeuPheGlyThrIleGlnSerIleGlyThrSerGlnValAsnLeuLys 862
2353 ...GACACGCACTATACGGTTACGCGCAACGCCACC..... 2385
863 GluAsnSerHisTrpHisLeuThrGlyAsnSerAsnValAsnGlnLeuAs 879
2386 ....CAAAACGGCAACCTCAGCCTCGTGGGCAATGCCCAA..... 2421
879 nLeuThrAsnGlyHisIleHisLeu.....AsnAlaGlnAsnAspAlaA 894
2422 .....GCAACATTTAATCAAGCCACATTAAACGGCAACACATCGGCT 2463
894 snLysValThrThrTyrAsnThrLeuThrValAsn.....SerLeu 907
2464 TCGGACAATGCTTCATTT.....AATCTAAG 2489
908 SerGlyAsnGlySerPheTyrTyrTrpValAspPheThrAsnAsnLysSe 924
2490 CAACAACGCCGTACAAAACGGCAGT..... 2514
924 rAsnLysValValValAsnLysSerAlaThrGlyAsnPheThrLeuGlnV 941
2515 .....CTGACGCTTTCCGAC 2529
941 alAlaAspLysThrGlyGluProAsnHisAsnGluLeuThrLeuPheAsp 957
2530 AACGCTAAGGCA.....AACGTAAGCCATTCCGCACTCAACGGCAA 2570
958 AlaSerAsnAlaThrArgAsnAsnLeuGluValThrLeuAlaAsnGlySe 974
2571 TGTCTCCCTAGCCGATAAGGCAGTATTCCATTTTGAAAACAGCCGCTTTA 2620
974 rVal.....AspArgGlyAlaTrpLysTyrLysLeuArgAsnValA 988
2621 CCGGAAAAATCAGCGGCGGCAAGGATACGGCATTACACTTAAAAGACAGC 2670
988 snGlyArgTyrAsp.....LeuTyrAsnPro 996
2671 GAATGGACGCTGCCGTCGGGCACGGAATTAGGCAATTTAAACCTTGACAA 2720
997 Glu.....ValGluLysArgAsnGlnThrValAspTh 1007
2721 CGCCACCATTACACTCAATTCCGCCTATCGACACGATGCGGCAGGCGCGC 2770
1007 rThrAsnIleThrThrProAsnAspIleGlnAlaAspAlaProSerAlaG 1024
2771 AAACCGGCAGT..... 2781
1024 lnSerAsnAsnGluGluIleAlaArgValGluThrProValProProPro 1040
2782 .....GCGGCAGATGCGCCGCGCCGCCGT.. 2805
1041 AlaProAlaThrGluSerAlaIleAlaSerGluGlnProGluThrArgPr 1057
2805 2805
1057 oAlaGluThrAlaGlnProAlaMetGluGluThrAsnThrAlaAsnSerT 1074


2806 .....TCG 2808
1074 hrGluThrAlaProLysSerAspThrAlaThrGlnThrGluAsnProAsn 1090
2809 CGCCGTTCCCTATTATCCGTTACG..... 2832
1091 SerGluSerValProSerGluThrThrGluLysValAlaGluAsnProPr 1107
2832 2832
1107 oGlnGluAsnGluThrValAlaLysAsnGluGlnGluAlaThrGluProT 1124
2833 .....CCGCCAACTTCGGCA 2847
1124 hrProGlnAsnGlyGluValAlaLysGluAspGlnProThrValGluAla 1140
2848 GAATCCCGTTTCAACACGCTGACGGTAAAC...GGCAAATTGAACGGTCA 2894
1141 AsnThrGlnThrAsnGluAlaThrGlnSerGluGlyLysThrGluGluTh 1157
2895 GGGAACATTCCGCTTTATGTCGGAA..... 2919
1157 rGlnThrAlaGluThrLysSerGluProThrGluSerValThrValSerG 1174
2920 .....CTCTTC 2925
1174 luAsnGlnProGluLysThrValSerGlnSerThrGluAspLysValVal 1190
2926 GGCTACCGCAGCGGCAAATTGAAGCTG..... 2952
1191 ValGluLysGluGluLysAlaLysValGluThrGluGluThrGlnLysAl 1207
2952 2952
1207 aProGlnValThrSerLysGluProProLysGlnAlaGluProAlaProG 1224
2953 .....GCGGAAAGTTCCGAAGGC 2970
1224 luGluValProThrAspThrAsnAlaGluGluAlaGlnAlaLeuGlnGln 1240
2971 ACTTACACCTTGGCTGTC..... 2988
1241 ThrGlnProThrThrValAlaAlaAlaGluThrThrSerProAsnSerLy 1257
2988 2988
1257 sProAlaGluGluThrGlnGlnProSerGluLysThrAsnAlaGluProV 1274
2989 .....AACAATACCGGCAACGAACCCGTAAGTCTCGAG 3021
1274 alThrProValValSerGluAsnThrAlaThrGlnProThrGluThrGlu 1290
3022 CAATTGACGGTAGTGGAAGGAAAAGACAACACACCGCTGTCCGAAAATCT 3071
1291 GluThrAlaLysValGlu.LysGluLysThrGlnGluValProGlnVal. 1306
3072 TAATTTCACCCTGCAAAACGAACACGTCGATGCCGGCGCATGGCGTTATC 3121
1307 .....AlaSerGlnGluSerPro.LysGlnGluGlnProAl 1318
3122 AGCTTATCCGCAAAGACGGCGAGTTCCGCCTGCATAATCCGGTCAAAGAA 3171
1318 aAlaLysProGlnAlaGlnThrLysProGlnAla.GluProAlaArgGlu 1334
3172 CAAGAGCTTTCCGACAAA..... 3189
1335 AsnValLeuThrThrLysAsnValGlyGluProGlnProGlnAlaGlnPr 1351
3190 .....CTCGGCAAGGCGGGAGAAACAGAGGCC...G 3217
1351 oGlnThrGlnSerThrAlaValProThrThrGlyGluThrAlaAlaAsnS 1368
3218 CCTTGACGGCAAAACAGGCACAACTTGCCGCCAAACAACAGGCGGAA... 3264

```
1368 erLysProAlaAlaLysProGlnAlaGlnAlaLysProGlnThrGluPro 1384

3265 ..........................AAAGACAACGCGCAAAGCCT 3284
                                  |||:::      ::::::
1385 AlaArgGluAsnValSerThrValAsnThrLysGluProGlnSerGlnTh 1401

3285 TGACGCGCTGATTGCGGCCGGGCGCAATGCCACCGAAAAGGCAGAAAGTG 3334
      :::|||   :::::::::   :::   |||   |||   :::::::::|
1401 rSerAlaThrValSerThrGluGlnProAlaLysGluThrSerSerAsnV 1418

3335 TTGCCGAACCGGCC.............................. 3348
     ||   :::||||||
1418 alGluGlnProAlaProGluAsnSerIleAsnThrGlySerAlaThrThr 1434

3348 ..... 3348

1435 MetThrGluThrAlaGluLysSerAspLysProGlnMetGluThrValTh 1451

3349 ..........CGGCAG.................GCAGGCGGGGAAAATG 3370
                |||||                 |||   ::::::   |
1451 rGluAsnAspArgGlnProGluAlaAsnThrValAlaAspAsnSerValA 1468

3371 CCGGCATTATGCAGGCGGAGGAAGAGAAAAAACGG............GTG 3408
     ||:::      :::::::::|||:::|||:::|||           |||
1468 laAsnAsnSerGluSerSerGluSerLysSerArgArgArgArgSerVal 1484

3409 CAGGCGGATAAAGACACCGCCTTGGCGAAACAGCGCGAAGCGGAAACCCG 3458
     :::      |||:::|||:::      :::          |||:::|||::
1485 SerGlnProLysGluThrSerAlaGluGluThrThrValAlaSerThrGl 1501

3459 GCCGGCTACC.................ACCGCCTTCCCCCGCGCCCGCC 3490
     :   :::|||                 |||      ||||||:::||||
1501 nGluThrThrValAspAsnSerValSerThrProLysProArgSerArgA 1518

3491 GCGCCCGCCGGGATTTGCCGCAACCGCAGCCCCAACCGCAACCCCAACCG 3540
     ||:::||||||::::::         :::   :::|||          |||
1518 rgThrArgArgSerValGlnThrAsnSerTyrGluProValGluLeuPro 1534

3541 CAG..... 3543

1535 ThrGluAsnAlaGluAsnAlaGluAsnValGlnSerGlyAsnAsnValAl 1551

3544 ..................CGCGACCTGATCAGCCGTTATGCCAATAGCG 3574
                       |||:::|||   |||:::   :::|||:::
1551 aAsnSerGlnProAlaLeuArgAsnLeuThrSerLysAsnThrAsnAlaV 1568

3575 GTTTGAGTGAATTTTCCGCC.............ACGCTCAACAGCGTT 3609
       |||||||:::      |||            :::|||||
1568 alLeuSerAsnAlaValAlaLysAlaGlnPheValAlaLeuAsnValGly 1584

3610 TTCGCCGTACAGGACGAATTGGACCGCGTG...TTTGCCGAAGACCGCCG 3656
        ||||||:::::::::::::::::::::::   :::   ::::::   ::
1585 LysAlaValSerGlnHisIleSerGlnLeuGluMetAsnAsnGluGlyGl 1601

3657 CAACGCCGTTTGGACAAGCGGCATCCGGGACACCAAACACTACCGTTCGC 3706
     :      ||||||   |||:::          :::|||:::|||   |||:
1601 nTyrAsnValTrpIleSerAsnThrSerMetAsnLysAsnTyrSerSerG 1618

3707 AAGATTTCCGCGCCTACCGCCAACAAACCGACCTGCGCCAAATCGGTATG 3756
     :::::::::|||   :::   :::::::::          |||:::|||
1618 luGlnTyrArgArgPheSerSerLysSerThrGlnThrGlnLeuGlyTrp 1634

3757 CAGAAAAACCTCGGCAGCGGG...CGCGTCGGCATCCTGTTTTCGCACAA 3803
     :::::::::::::::::::::   :::::::|||   :::|||::::::
1635 AspGlnThrIleSerAsnAsnValGlnLeuGlyGlyValPheThrTyrVa 1651

3804 CCGGACCGGAAACACCTTCGACGACGGCATCGGCAACTCGGCACGGCTTG 3853
      |||:::::::|||:::||||||   :::   ::::::::::::
1651 lArgAsnSerAsnAsnPheAspLysAlaSerSerLysAsnThrLeuAlaG 1668

3854 CCCACGGTGCCGTTTTCGGGCAATACGGCATCGGCAGGTTCGACATCGGC 3903
           :::               |||:::   :::::::::   :::|||

1668 lnValAsnPheTyrSerLysTyrTyrAlaAspAsnHisTrpTyrLeuGly 1684

3904 ATCAGCGCGGGCGCGGGTTTTAGTAGCGGCAGCCTTTCAGACGGCATCAG 3953
     |||:::   |||   |||      :::::::::|||:::   :::   ::
1685 IleAspLeuGlyTyrGlyLysPheGlnSerAsnLeuGlnThrAsnAsnAs 1701

3954 AGGCAAAATCCGCCGCCGCGTGCTGCATTACGGCATTCAGGCAAGATACC 4003
     ::::|||:::   |||::::::   :::   |||:::   |||
1701 nAlaLysPheAlaArgHisThrAlaGlnIleGlyLeuThrAlaGlyLysA 1718

4004 GCGCAGGTTTCGGCGGATTCGGCATCGAACCGCACATCGGCGCAACGCGC 4053
          :::::::|||:::|||:::::::::|||   :::|||:::
1718 laPheAsnLeuGlyAsnPheAlaValLysProThrValGlyValArgTyr 1734

4054 TATTTCGTCCAAAAAGCGGATTACCGATACGAAAACGTCAATATCGCCAC 4103
        ::::::::::::|||||||:::          :::   :::|||   ::
1735 SerTyrLeuSerAsnAlaAspPheAlaLeuAlaGlnAspArgIleLysVa 1751

4104 CCCGGGCCTTGCATTCAACCGCTACCGCGCGGGCATTAAGGCAGATTATT 4153
     :      ::::::   :::          |||   :::      :::|||:
1751 lAsnProIleSerValLysThrAlaPheAlaGlnValAspLeuSerTyrT 1768

4154 CATTCAAACCGGCGCAACACATTTCCATCACGCCTTATTTGAGCCTGTCC 4203
     :::::      ::::::   :::||||||||||||   ||||||
1768 hrTyrHisLeuGlyGlu...PheSerIleThrProIleLeuSerAlaArg 1783

4204 TATACCGATGCCGCTTCCGGCAAAGTCCGAACGCGCGTCAATACCGCCGT 4253
     |||   ||||||   :::|||:::   :::   :::|||::::::
1784 Tyr...AspAlaAsnGlnGlyAsnGlyLysIleAsnValSerValTyrAs 1799

4254 ATTGGCGCAGGATTTCGGCAAAACCCGCAGTGCGGAATGGGGCGTAAACG 4303
      :::|||   :::      :::   ::::::   :::   |||::::::
1799 pPheAlaTyrAsnValGluAsnGlnGlnGlnTyrAsnAlaGlyLeuLysL 1816

4304 CCGAAATCAAAGGTTTCACGCTGTCCCTCCACGCTGCCGCCGCCAAGGGG 4353
       :::      :::      |||||||||   ::::::   :::|||:::
1816 euLysTyrHisAsnValLysLeuSerLeuIleGlyGlyLeuThrLysAla 1832

4354 CCGCAATTGGAAGCGCAGCACAGCGCGGGCATCAAATTAGGCTAC 4398
       |||   |||   |||   :::|||   :::||||||::::::
1833 LysGlnAlaGluLysGlnLysThrAlaGluValLysLeuSerPhe 1847
```

seq_name: pir2:A81937

seq_documentation_block:
IgA-specific metalloendopeptidase (EC 3.4.24.13) NMA0905 [imported] - Neisseria menin
N;Alternate names: IgA proteinase; IgA1 proteinase (EC 3.4.21.7) [misnomer]; immunogl
C;Species: Neisseria meningitidis
A;Variety: strain Z2491 serogroup A; strain HF117; strain HF159; strain SM1027
C;Date: 05-May-2000 #sequence_revision 05-May-2000 #text_change 02-Feb-2001
C;Accession: A81937; S61317; S61318; S61321
R;Parkhill, J.; Achtman, M.; James, K.D.; Bentley, S.D.; Churcher, C.; Klee, S.R.; Mo
; Holroyd, S.; Jagels, K.; Leather, S.; Moule, S.; Mungall, K.; Quail, M.A.; Rajandre
Nature 404, 502-506, 2000
A;Title: Complete DNA sequence of a serogroup A strain of Neisseria menigitidis Z2491
A;Reference number: A81775; MUID:20222556
A;Accession: A81937
A;Molecule type: DNA
A;Residues: 1-1773 <PAR>
A;Cross-references: GB:AL162754; GB:AL157959; NID:g7379424; PIDN:CAB84182.1; PID:g737
A;Experimental source: serogroup A, strain Z2491
R;Lomholt, H.; Poulsen, K.; Mogens, K.
Mol. Microbiol. 15, 495-506, 1995
A;Title: Comparative characterization of the iga gene encoding IgA1 protease in Neiss
A;Reference number: S61314; MUID:95302961
A;Accession: S61317
A;Status: nucleic acid sequence not shown
A;Molecule type: DNA
A;Residues: 53-548 <LOM>
A;Cross-references: EMBL:X82470; NID:g732854; PIDN:CAA57853.1; PID:g732855
A;Experimental source: strain HF117
A;Accession: S61318
A;Status: nucleic acid sequence not shown

```
A;Molecule type: DNA
A;Residues: 53-548 <LOW>
A;Cross-references: EMBL:X82471; NID:g732858; PIDN:CAA57854.1; PID:g732859
A;Experimental source: strain HF159
A;Accession: S61321
A;Status: nucleic acid sequence not shown
A;Molecule type: DNA
A;Residues: 53-548 <LOK>
A;Cross-references: EMBL:X82472; NID:g732852; PIDN:CAA57855.1; PID:g732853
A;Experimental source: strain SM1027
C;Genetics:
A;Gene: iga; NMA0905
C;Superfamily: IgA-specific metalloendopeptidase
C;Keywords: hydrolase; metalloproteinase


alignment_scores:
           Quality: 1515.00              Length:   1958
             Ratio:  1.647                Gaps:     53
 Percent Similarity:  46.987   Percent Identity: 24.974

alignment_block:
 US-09-303-518D-653 x A81937   ..

 Align seg 1/1  to: A81937  from: 1  to: 1773

      49 AAAACCGGCCGCATCCGCTTCTCGCCCGCTTACTTAGCCATATGCCTGTC 98
         ||||||   |||:::::::::::::         |||:::|||   |||::
       '2 LysThrLysArgPheLysIleAsnAlaIleSerLeuSerIlePheLeuAl 18

      99 GTTCGGCATTCTGCCCCAAGCCCGGGCGGGACACACTTATTTCGGCATCA 148
         ::::::::::   |||   ::::::|||:::   :::          ::::
      18 aTyrAlaLeuThrProTyrSerGluAlaAlaLeuValArgAspAspValA 35

     149 ACTACCAATACTATCGCGACTTTGCCGAAAATAAAGGCAAGTTTGCAGTC 198
         ::|||||||   :::|||||||||||||||||||||||||||||||   |||
      35 spTyrGlnIlePheArgAspPheAlaGluAsnLysGlyLysPhePheVal 51

     199 GGGGCGAAAGATATTGAGGTTTACAACAAAAAAGGGGAGTTGGTCGGCAA 248
         ||||||   |||:::::::|||   ||||||:::|||:::   :::|||::
      52 GlyAlaThrAspLeuSerValLysAsnLysGlnGlyGlnAsnIleGlyAs 68

     249 ATCGATGACGAAAGCCCCGATGATTGATTTTTCTGTGGTATCGCGTAAC. 297
         ::::::::::::::::|||||||||||||||||||||:::::   |||
      68 nAlaLeuSerAsnValProMetIleAspPheSerValAlaAspValAsnA 85

     298 ..GGCGTGGCGGCATTGGCGGGCGATCAATATATTGTGAGCGTGGCACAT 345
             :::   ::::::         ||||||   |||||||||   |||
      85 rgArgThrLeuThrValIleAspProGlnTyrAlaValSerValLysHis 101

     346 AACGGCGGC.............TATAACAATGTTGATTTTGGTGCGGA 380
             |||               |||::::::   :::       ::
     102 ValLysGlyAspGluIleSerTyrTyrGlyHisHisAsnGlyHisLeuAs 118

     381 GGGAAGCAATCCCGATCAGCACCGCTTTTCTTACCAAATTGTGAAAAGAA 430
         :   ||||||   ::::::         :::|||:::   |||   :::|
     118 pValSerAsnAspGluAsn.........GluTyrArgSerValAlaGlnA 132

     431 ATAATTATAAAGCAGGGACTAACGGCCATCCTTATGGCGGC......... 471
         ||:::|||:::   :::   |||   |||   :::|||:::
     132 snAspTyrGluProAsnLysAsnTrpHis...HisGlyAsnGlnGlyArg 147

     472 ......GATTATCATATGCCGCGTTTGCACAAATTTGTCACAGATGCAGA 515
               ||||||:::|||   ||||||:::||||||||||||:::::::
     148 LeuGluAspTyrAsnMetAlaArgLeuAsnLysPheValThrGluValAl 164

     516 ACCTGTTGAGATGACCAGTTATATGGATGGGTGGAAA...TACGCTGATT 562
          |||:::      ||||||         |||   :::   |||   |||
     164 aProIleAlaProThrSerAlaGlyGlyGlyValGluThrTyrLysAspL 181

     563 TAAATAAATACCCTGATCGTGTTCGAATCGGAGCAGGCAGACAATATTGG 612
           |||::::::   :::   ||||||:::|||||||||   |||:::
     181 ysAsnArgPheSerGluPheValArgValGlyAlaGlyThrGlnPhe... 196

     613 CGGTCTGATGAAGACGAACCCAATAACCGCGAAAGTTCATATCATATTGC 662
                       |||   |||:::|||   :::       ::::::::
     197 ..............GluTyrAsnSerArgTyrAsnMetThrGluLeuSe 208

     663 AAGCGCATATTCTTGGCTCGTCGGTGGCAATACCTTT............. 699
         :   ||||||   :::   ::::::|||:::   :::
     208 rArgAlaTyrArgTyrAlaIleAlaGlyThrProTyrGlnAspValAsnV 225

     700 ..GCACAAAATGGATCAGGTGGTGGCACAGTCAACTTAGGTAGCGAAAAA 747
           ::::::|||   :::       |||   :::::::::|||:::::::::
     225 alThrSerAsnLeuAsnGlnGluGlyLeuIleGlyPheGlyAspAsnSer 241

     748 ATTAAACATAGCCCA.......................TATGGTTTTTT 773
               |||||||||                           :::   ||
     242 LysHisHisSerProGluLysLeuLysGluValLeuSerGlnAsnAlaLe 258

     774 ACCAACAGGAGGCTCATTTGGCGACAGTGGCTCACCAATGTTTATCTATG 823
         |   :::   :::   :::|||||||||||||||||||:::|||   ||||
     258 uThrAsnTyrAlaValLeuGlyAspSerGlySerProLeuPheAlaTyrA 275

     824 ATGCCCAAAAGCAAAAGTGGTTAATTAATGGGGTATTGCAAACAGGCAAC 873
         ||   |||:::::::::|||::::::   |||:::   :::
     275 spLysGlnGluLysArgTrpValPheLeuGlyAlaTyrAspTyrTrpAla 291

     874 CCCTATATAGGAAAAAGCAATGGCTTCCAGCTAGTTCGTAAAGATTGG.. 921
           |||       ::::::|||::::::|||           :::|||
     292 GlyTyr......GlnLysAsnSerTrpGln............GluTrpAs 302

     922 .TTCTATGATGAAATCTTTGCTGGAGATACCCATTCAGTATTCTACGAAC 970
          :::|||   :::   ||||||   :::       :::
     302 nIleTyrLysLysGluPheAlaAspGluIleLysGln.............. 314

     971 CACATCAAAATGGGAAATACTTTTTTAACGACAATAATAATGGCGCAGGA 1020
                                     :::||||||          ||||||
     315 ........................ArgAspAsn.........AlaGly 319

    1021 AAAATCGATGCCAAACATAAACACTATTCTCTACCTTATAGATTAAAAAC 1070
           |||   :::       :::|||:::                  |||||
     320 ThrIleLysGlyTyrGlyGluHisHis..............TrpLysTh 331

    1071 ACGAACCGTTCAATTGTTTAATGTTTCTTTATCCGAGACAGCAAGAGAAC 1120
         |       :::          |||       :::::::::||||||   :::
     331 rThrGlyThr.........AsnSerHisIleGlySerThrAlaValArgL 345

    1121 CTGTTTATCATGCTGCAGGTGGGGTCAACAGTTATCGACCCAGACTGAAT 1170
           :::   :::       |||:::                      |||
     345 euAlaGlyAsnGluArgGlyAla.......................Asn 353

    1171 AATGGAGAAAATATTTCCTTTATTGACAAAGGAAAAGGTGAATTGATACT 1220
         ||||||:::|||::::::|||       :::::::::|||   |||:::||
     354 AsnGlyGlnAsnValThrPhe......GluAsnAsnGlyThrLeuValLe 368

    1221 TACCAGCAACATCAACCAAGGCGCGGGCGGTTTGTATTTTGAGGGTAATT 1270
         |   :::|||||||||||||||||||||||||||:::|||:::|||::::
     368 uAspGlnAsnIleAsnGlnGlyAlaGlyGlyLeuPhePheLysGlyAspT 385

    1271 TTACGGTCTCGCCTAAAAACAACGAA...ACGTGGCAAGGCGCGGGCGTT 1317
         ::||||||:::       ||||||      ||||||   |||||||||:::
     385 yrThrValLysGlyAlaAsnAsnGlyIleThrTrpLeuGlyAlaGlyIle 401

    1318 CATATCAGTGATGGCAGTACCGTTACTTGGAAAGTAAACGGCGTGGCAAA 1367
         ::::::||||||:::   |||:::|||:::|||::::::       ::
     402 AspValAlaAspGlyLysLysValValTrpGlnValLysAsnProAsnGl 418

    1368 CGACCGCCTGTCCAAAATCGGCAAAGGCACGCTGCTGGTTCAAGCCAAAG 1417
         :|||||||||:::|||||||||||||||||||||   :::::::::   |
     418 yAspArgLeuAlaLysIleGlyLysGlyThrLeuGluIleAsnGlyThrG 435

    1418 GGGAAAACCAAGGCTCGGTCAGCGTGGGCGACGGTAAAGTCATCTTAGAT 1467
         ||   |||||||||:::::::::||||||||||   ||||||||||:::
     435 lyValAsnGlnGlyGlnLeuLysValGlyAspGlyThrValIleLeuAsn 451
```



```
 921 LeuPheAspAlaSerSerValArgAspArgSer...HisLeuSerValSe 936

3081 CCTGCAAAACGAACACGTCGATGCCGGCGCATGGCGTTATCAGCTTATCC 3130
     :|||   |||:::|||||||||   ||||||   ||||||   :::
 936 rLeuAlaAsnHisHisValAspLeuGlyAlaLeuArgTyrThrIleLysT 953

3131 GCAAAGACGGCGAGTTCCGCCTGCATAATCCG..... 3162
       ::::::|||      ||||||:::||||||
 953 hrGluAsnGlyIleThrArgLeuTyrAsnProTyrAlaGluAsnArgArg 969

3162 ..... 3162

 970 ArgValLysProAlaProSerProAlaThrAsnThrAlaSerGlnAlaGl 986

3162 ..... 3162

 986 nThrAspSerAlaGlnIleAlaLysProGlnAsnIleValValAlaProP 1003

3162 ..... 3162

1003 roSerProGlnAlaAsnGlnAlaGluGluAlaLysArgGlnGlnAlaLys 1019

3163 .....GTCAAAGAACAA..... 3174
           ||||||:::|||
1020 AlaGluGlnValLysArgGlnGlnAlaGluAlaGluArgLysSerAlaGl 1036

3174 ..... 3174

1036 uLeuAlaLysGlnLysAlaGluAlaGluArgGluAlaArgGluLeuAlaT 1053

3174 ..... 3174

1053 hrArgGlnLysAlaGluGlnGluArgSerSerAlaGluLeuAlaArgArg 1069

3175 .....GAGCTTTCCGACAAA..... 3189
                          |||||||||   |||
1070 HisGluLysGluArgGluAlaAlaGluLeuSerAlaLysGlnLysValGl 1086

3189 ..... 3189

1086 uAlaGluArgGluAlaGlnAlaLeuAlaValArgArgLysAlaGluAlaG 1103

3190 .....CTCGGCAAGGCGGGAGAAACAGAG 3213
                          |||::::::      |||   |||
1103 luGluAlaLysArgGlnAlaAlaGluLeuAlaArgArgHisGluLysGlu 1119

3214 .....GCCGCCTTGACGGCAAAACAG..... 3234
           |||   |||:::|||||||||
1120 ArgGluAlaAlaGluLeuSerAlaLysGlnArgValGlyGluGluGluAr 1136

3234 ..... 3234

1136 gArgGlnThrAlaGlnSerGlnProGlnArgArgLysArgArgAlaAlaP 1153

3234 ..... 3234

1153 roGlnAspTyrMetAlaAlaSerGlnAspArgProLysArgArgGlyHis 1169

3234 ..... 3234

1170 ArgSerValGlnGlnAsnAsnValGluIleAlaGlnAlaGlnAlaGluLe 1186

3235 .....GCACAACTTGCCGCCAAAC 3253
                          |||:::|||   |||||||
1186 uAlaArgArgGlnGlnGluGluArgLysAlaAlaGluLeuLeuAlaLysG 1203

3254 AACAGGCGGAAAAAGAC...AACGCGCAAAGCCTTGACGCGCTG..... 3294
     ||:::||||||   :::   :::|||||:::|||   |||
1203 lnArgAlaGluAlaGluArgGluAlaGlnAlaLeuAlaAlaArgArgLys 1219

3294 ..... 3294

1220 AlaGluAlaGluGluAlaLysArgGlnAlaAlaGluLeuAlaHisArgGl 1236

3295 .....ATTGCGGCCGGGCGCAATGCCACCG 3319
                          ::::::|||:::::::::|||:::
1236 nGluAlaGluArgLysAlaAlaGluLeuSerAlaAsnGlnLysAlaAlaA 1253

3320 AAAAGGCAGAAAGTGTTGCC..... 3339
       :::|||::::::::::|||
1253 laGluAlaGlnAlaLeuAlaAlaArgGlnGlnLysAlaLeuAlaArgGln 1269

3340 ...GAACCGGCCCGGCAGGCAGGC..... 3360
        |||   ||||||:::|||:::
1270 GlnGluGluAlaArgLysAlaAlaGluLeuAlaValLysGlnLysAlaGl 1286

3361 ................GGGGAAAATGCCGGCATTATGCAGGCGGAGGAAG 3394
                       :::|||   |||           |||   |||
1286 uThrGluArgLysThrAlaGluLeuAlaLysGlnArgAlaAlaAlaGluA 1303

3395 AGAAAAAACGGGTGCAGGCGGATAAAGACACCGCC...TTGGCGAAACAG 3441
          ||||||   |||      ::::::||||||   ||||||::::::
1303 laAlaLysArgGlnGlnGluAlaArgGlnThrAlaGluLeuAlaArgArg 1319

3442 CGCGAAGCGGAAACCCGGCCGGCTACCACCGCCTTC..... 3477
     :::|||||||||   :::   |||      :::
1320 GlnGluAlaGluArgGlnAlaAlaGluLeuSerAlaLysGlnLysAlaGl 1336

3477 ..... 3477

1336 uThrAspArgGluAlaAlaGluSerAlaLysArgLysAlaGluGluGluG 1353

3477 ..... 3477

1353 luHisArgGlnAlaAlaGlnSerGlnProGlnArgArgLysArgArgAla 1369

3478 .....CCCCGCGCCCGCCG 3491
                          |||:::   |||
1370 AlaProGlnAspTyrMetAlaAlaSerGlnAsnArgProLysArgArgGl 1386

3492 CGCCCGCCGGGATTTGCCGCAACCG..... 3516
        |||      ||||||   |||
1386 yArgArgSerThrLeuProAlaProProSerProSerPheAspSerSerA 1403

3516 ..... 3516

1403 laTyrAlaAlaProArgAlaLeuHisAsnProAspTrpTyrGluAsnAsp 1419

3516 ..... 3516

1420 TyrGluGluIleProLeuAspAlaLeuGluAspGluAsnValSerGluSe 1436

3516 ..... 3516

1436 rValAspThrSerAspLysGlnProGlnAspAsnThrGluLeuHisGluL 1453

3516 ..... 3516

1453 ysValGluThrValSerLeuGlnProArgAlaAlaGlnProArgAlaGln 1469

3517 .....CAGCCCCAACCGCAACCCCAACCGCAGCGCGACCTGATCAG 3557
               |||||||||   |||          |||   |||   :::||
1470 AlaAlaThrGlnProGlnAlaGlnAlaAlaAlaGlnAlaAspAlaValSe 1486

3558 CCGTTATGCCAATAGCGGTTTGAGTGAATTTTCCGCCACGCTCAACAGCG 3607
     |      ::::||||||:::||||||:::      |||:::   :::|||:
1486 rThrAsnThrAsnSerAlaLeuSerAspAlaMetAlaSerThrGlnSerI 1503

3608 TTTTCGCC...GTACAGGACGAATTGGACCGCGTGTTTGCCGAAGACCGC 3654
     :::::      :::      :::|||   |||   :::|||:::
1503 leLeuLeuAspThrGlyAlaSerLeuThrArgHisIleAlaGlnLysSer 1519

3655 CGC..............AACGCCGTTTGGACAAGCGGCATCCGGGACAC 3689
     |||               |||:::||||||   |||:::|||
1520 ArgAlaAspAlaGluLysAsnSerValTrpMetSerAsnIleGlyTyrGl 1536
```

```
  3690 CAAACACTACCGTTCGCAAGATTTCCGCGCCTACCGCCAACAAACCGACC 3739
       :::   |||   |||   ::::::|||   :::   ::::::
  1536 yArgAspTyrAlaSerAlaGlnTyrArgArgPheSerSerLysArgThrG 1553

  3740 TGCGCCAAATCGGTATGCAGAAAAACCTCGGCAGCGGG...CGCGTCGGC 3786
       |||||||||:::::::::::::|||::::::::::   ::::::|||
  1553 lnThrGlnIleGlyIleAspArgSerLeuSerGluAsnMetGlnIleGly 1569

  3787 ATCCTGTTTTCGCACAACCGGACCGGAAACACCTTCGACGACGGCATCGG 3836
       ::::::::::::::   :::   :::|||||||||::::::   ||
  1570 GlyValLeuThrTyrSerAspSerGlnHisThrPheAspGlnAlaSerGl 1586

  3837 CAACTCGGCACGGCTTGCCCACGGTGCCGTTTTCGGGCAATACGGCATCG 3886
       |:::::::::   ::::::::::::   ::::::|||:::|||   ::::
  1586 yLysAsnThr...PheValGlnAlaAsnLeuTyrGlyLysTyrTyrLeuA 1602

  3887 GC...AGGTTCGACATCGGCATCAGCGCGGGCGCGGGTTTTAGTAGCGGC 3933
       ::      :::   ::::::   :::   |||||||||         :::
  1602 snAspAlaTrpTyrMetAlaGlyAspIleGlyAlaGlySerLeuArgSer 1618

  3934 AGCCTTTCAGACGGCATCAGAGGCAAAATCCGCCGCCGCGTGCTGCATTA 3983
       |||:::         ::::::::::::::|||      ::::::
  1619 ArgLeuGlnThrGlnGlnLysAlaAsnPheAsnArgThrSerIleGlnTh 1635

  3984 CGGCATTCAGGCAAGATACCGCGCAGGTTTCGGCGGATTCGGCATCGAAC 4033
       |||:::                  ::::::   |||   |||   |
  1635 rGlyLeuThrLeuGlyAsnThrLeuLysIleAsnGlnPheGluIleValP 1652

  4034 CGCACATCGGCGCAACGCGCTATTTCGTCCAAAAAGCGGATTACCGATAC 4083
       ||      |||            :::::::::|||||||||:::
  1652 roSerAlaGlyIleArgTyrSerArgLeuSerSerAlaAspTyrLysLeu 1668

  4084 GAAAACGTCAATATCGCCACCCCGGGCCTTGCATTCAACCGCTACCGCGC 4133
       |||   ::::::   :::   :::::::::   :::          ||
  1669 GlyAsnAspSerValLysValSerSerMetSerValLysThrLeuThrAl 1685

  4134 GGGCATTAAGGCAGATTATTCATTCAAACCGGCGCAACACATTTCCATCA 4183
       ||||:::         |||   ||||||   :::   ::::::::::::
  1685 aGlyLeuAspPheAlaTyrArgPheLys...ValGlyAsnLeuThrValL 1701

  4184 CGCCTTATTTGAGCCTGTCCTAT...ACCGATGCCGCTTCCGGCAAAGTC 4230
       |||   ||||||   :::|||   ::::::   :::::::|||   |||
  1701 ysProLeuLeuSerAlaAlaTyrPheAlaAsnTyrGlyLysGlyGlyVal 1717

  4231 CGAACGCGCGTCAATACCGCCGTATTGGCGCAGGATTTCGGCAAAACCCG 4280
       ::::::      |||:::   |||         |||      :::
  1718 AsnValGlyGlyAsnSerPheValTyrLysAlaAspAsnGlnGlnGlnTy 1734

  4281 CAGTGCGGAATGGGGCGTAAACGCCGAAATCAAAGGTTTCACGCTGTCCC 4330
       ||||||      |||:::            ::::::   ||||||::::
  1734 rSerAla......GlyAlaAlaLeuLeuTyrArgAsnValThrLeuAsnV 1749

  4331 TCCACGCTGCCGCCGCCAAGGGGCCGCAATTGGAAGCGCAGCACAGCGCG 4380
       ::::::::::   :::||||||   |||||||||   |||   |||:::
  1749 alAsnGlySerIleThrLysGlyLysGlnLeuGluLysGlnLysSerGly 1765

  4381 GGCATCAAATTAGGCTACCGCTGG 4404
       ||||||:::      |||:::
  1766 GlnIleLysIleGlnIleArgPhe 1773

seq_name: pir2:C81169

seq_documentation_block:
IgA-specific metalloendopeptidase (EC 3.4.24.13) NMB0700 [imported] - Neisseria meningit
C;Species: Neisseria meningitidis
A;Variety: strain MC58 serogroup B; strain 81139
C;Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 19-Jan-2001
C;Accession: C81169; S61326
R;Tettelin, H.; Saunders, N.J.; Heidelberg, J.; Jeffries, A.C.; Nelson, K.E.; Eisen, J.A
Hickey, E.K.; Haft, D.H.; Salzberg, S.L.; White, O.; Fleischmann, R.D.; Dougherty, B.A.;
ri, H.; Qin, H.; Vamathevan, J.; Gill, J.; Scarlato, V.; Masignani, V.; Pizza, M.
Science 287, 1809-1815, 2000
A;Authors: Grandi, G.; Sun, L.; Smith, H.O.; Fraser, C.M.; Moxon, E.R.; Rappuoli, R.; Ve
A;Title: Complete genome sequence of Neisseria meningitidis serogroup B strain MC58.
A;Reference number: A81000; MUID:20175755
A;Accession: C81169
A;Molecule type: DNA
A;Residues: 1-1815 <TET>
A;Cross-references: GB:AE002424; GB:AE002098; NID:g7225923; PIDN:AAF41117.1; PID:g722
A;Experimental source: serogroup B, strain MC58
R;Lomholt, H.; Poulsen, K.; Mogens, K.
Mol. Microbiol. 15, 495-506, 1995
A;Title: Comparative characterization of the iga gene encoding IgA1 protease in Neiss
A;Reference number: S61314; MUID:95302961
A;Accession: S61326
A;Status: preliminary; nucleic acid sequence not shown
A;Molecule type: DNA
A;Residues: 53-548 <LOM>
A;Cross-references: EMBL:X82477; NID:g732856; PIDN:CAA57860.1; PID:g732857
C;Genetics:
A;Gene: NMB0700
C;Superfamily: IgA-specific metalloendopeptidase
C;Keywords: hydrolase; metalloproteinase


alignment_scores:
            Quality: 1486.00            Length:   2000
              Ratio:  1.613              Gaps:     56
 Percent Similarity:  46.050   Percent Identity: 24.600

alignment_block:
 US-09-303-518D-653 x C81169   ..

 Align seg 1/1  to: C81169  from: 1  to: 1815

     49 AAAACCGGCCGCATCCGCTTCTCGCCCGCTTACTTAGCCATATGCCTGTC 98
        ||||||   |||:::::::::::::         |||:::|||   |||::
      2 LysThrLysArgPheLysIleAsnAlaIleSerLeuSerIlePheLeuAl 18

     99 GTTCGGCATTCTGCCCCAAGCCCGGGCGGGACACACTTATTTCGGCATCA 148
        ::::::::::   |||   ::::::|||:::   :::         ::::
     18 aTyrAlaLeuThrProTyrSerGluAlaAlaLeuValArgAspAspValA 35

    149 ACTACCAATACTATCGCGACTTTGCCGAAAATAAAGGCAAGTTTGCAGTC 198
        ::|||||   :::||||||||||||||||||||||||||||||   |||
     35 spTyrGlnIlePheArgAspPheAlaGluAsnLysGlyLysPhePheVal 51

    199 GGGGCGAAAGATATTGAGGTTTACAACAAAAAAGGGGAGTTGGTCGGCAA 248
        ||||||   |||::::::|||   ||||||:::|||:::   :::|||::
     52 GlyAlaThrAspLeuSerValLysAsnLysGlnGlyGlnAsnIleGlyAs 68

    249 ATCGATGACGAAAGCCCCGATGATTGATTTTTCTGTGGTATCGCGTAAC. 297
        ::::::::::::::::|||||||||||||||||||||||::::::   |||
     68 nAlaLeuSerAsnValProMetIleAspPheSerValAlaAspValAsnA 85

    298 ..GGCGTGGCGGCATTGGCGGGCGATCAATATATTGTGAGCGTGGCACAT 345
           :::   ::::::         ||||||   |||||||||   |||
     85 rgArgThrLeuThrValIleAspProGlnTyrAlaValSerValLysHis 101

    346 AACGGCGGC..............TATAACAATGTTGATTTTGGTGCGGA 380
           |||               |||::::::   :::         ::
    102 ValLysGlyAspGluIleSerTyrTyrGlyHisHisAsnGlyHisLeuAs 118

    381 GGGAAGCAATCCCGATCAGCACCGCTTTTCTTACCAAATTGTGAAAAGAA 430
        :   ||||||   ::::::         :::|||:::   |||   :::|
    118 pValSerAsnAspGluAsn.........GluTyrArgSerValAlaGlnA 132

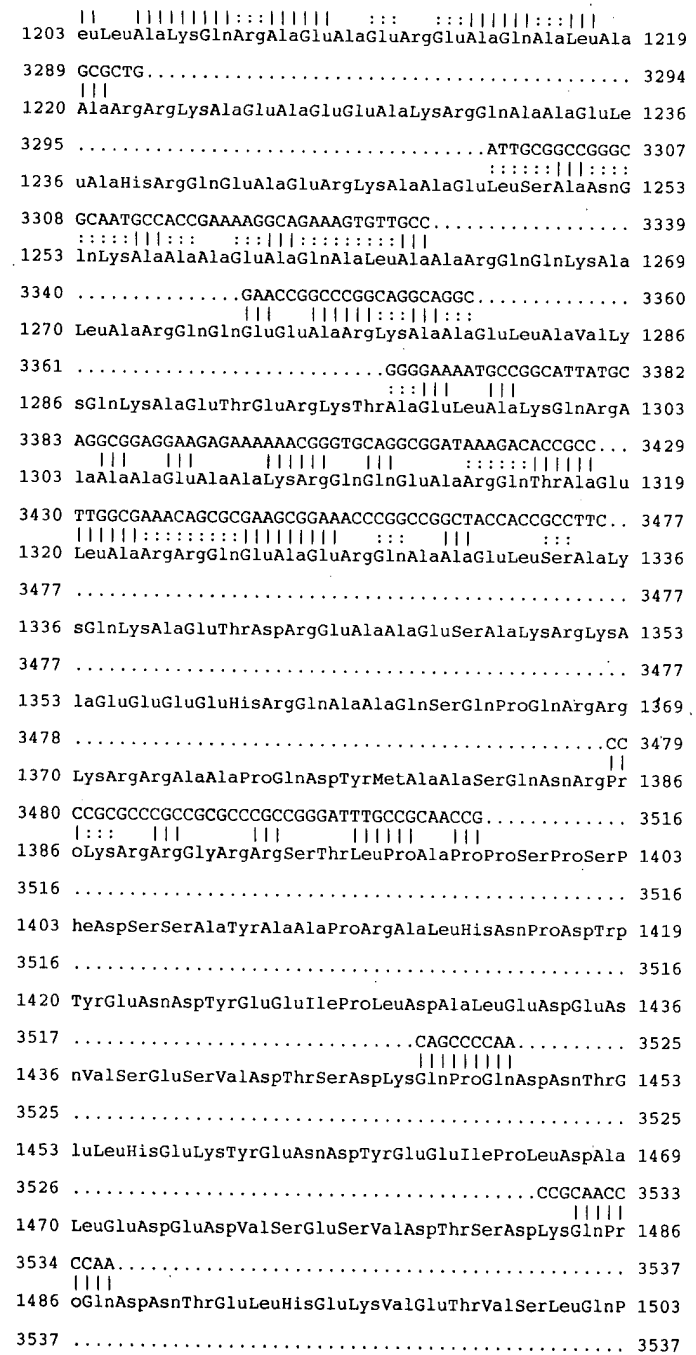
    431 ATAATTATAAAGCAGGGACTAACGGCCATCCTTATGGCGGC......... 471
        ||:::|||:::   :::   |||   |||   :::|||:::
    132 snAspTyrGluProAsnLysAsnTrpHis...HisGlyAsnGlnGlyArg 147

    472 ......GATTATCATATGCCGCGTTTGCACAAATTTGTCACAGATGCAGA 515
           ||||||:::|||   ||||||:::||||||||||||:::::
    148 LeuGluAspTyrAsnMetAlaArgLeuAsnLysPheValThrGluValAl 164

    516 ACCTGTTGAGATGACCAGTTATATGGATGGGTGGAAA...TACGCTGATT 562
```






1203 euLeuAlaLysGlnArgAlaGluAlaGluArgGluAlaGlnAlaLeuAla 1219
3289 GCGCTG 3294
1220 AlaArgArgLysAlaGluAlaGluGluAlaLysArgGlnAlaAlaGluLe 1236
3295 ATTGCGGCCGGGC 3307
1236 uAlaHisArgGlnGluAlaGluArgLysAlaAlaGluLeuSerAlaAsnG 1253
3308 GCAATGCCACCGAAAAGGCAGAAAGTGTTGCC 3339
1253 lnLysAlaAlaAlaGluAlaGlnAlaLeuAlaAlaArgGlnGlnLysAla 1269
3340 GAACCGGCCCGGCAGGCAGGC 3360
1270 LeuAlaArgGlnGlnGluGluAlaArgLysAlaAlaGluLeuAlaValLy 1286
3361 GGGGAAAATGCCGGCATTATGC 3382
1286 sGlnLysAlaGluThrGluArgLysThrAlaGluLeuAlaLysGlnArgA 1303
3383 AGGCGGAGGAAGAGAAAAAACGGGTGCAGGCGGATAAAGACACCGCC 3429
1303 laAlaAlaGluAlaAlaLysArgGlnGlnGluAlaArgGlnThrAlaGlu 1319
3430 TTGGCGAAACAGCGCGAAGCGGAAACCCGGCCGGCTACCACCGCCTTC 3477
1320 LeuAlaArgArgGlnGluAlaGluArgGlnAlaAlaGluLeuSerAlaLy 1336
3477 3477
1336 sGlnLysAlaGluThrAspArgGluAlaAlaGluSerAlaLysArgLysA 1353
3477 3477
1353 laGluGluGluGluHisArgGlnAlaAlaGlnSerGlnProGlnArgArg 1369
3478 CC 3479
1370 LysArgArgAlaAlaProGlnAspTyrMetAlaAlaSerGlnAsnArgPr 1386
3480 CCGCGCCCGCCGCGCCCGCCGGGATTTGCCGCAACCG 3516
1386 oLysArgArgGlyArgArgSerThrLeuProAlaProProSerProSerP 1403
3516 3516
1403 heAspSerSerAlaTyrAlaAlaProArgAlaLeuHisAsnProAspTrp 1419
3516 3516
1420 TyrGluAsnAspTyrGluGluIleProLeuAspAlaLeuGluAspGluAs 1436
3517 CAGCCCCAA 3525
1436 nValSerGluSerValAspThrSerAspLysGlnProGlnAspAsnThrG 1453
3525 3525
1453 luLeuHisGluLysTyrGluAsnAspTyrGluGluIleProLeuAspAla 1469
3526 CCGCAACC 3533
1470 LeuGluAspGluAspValSerGluSerValAspThrSerAspLysGlnPr 1486
3534 CCAA 3537
1486 oGlnAspAsnThrGluLeuHisGluLysValGluThrValSerLeuGlnP 1503
3537 3537


1503 roArgAlaAlaGlnProArgAlaGlnAlaAlaThrGlnLeuGlnAlaGln 1519
3538 .....CCGCAGCGCGACCTGATCAGCCGTTATGCCAATAGCGGTTTGAG 3581
1520 AlaAlaAlaGlnAlaAspAlaValSerThrAsnThrAsnSerAlaLeuSe 1536
3582 TGAATTTTCCGCCACGCTCAACAGCGTTTTCGCC...GTACAGGACGAAT 3628
1536 rAspAlaMetAlaSerThrGlnSerIleLeuLeuAspThrGlyAlaSerL 1553
3629 TGGACCGCGTGTTTGCCGAAGACCGCCGC.....AACGCC 3663
1553 euThrArgHisIleAlaGlnLysSerArgAlaAspAlaGluLysAsnSer 1569
3664 GTTTGGACAAGCGGCATCCGGGACACCAAACACTACCGTTCGCAAGATTT 3713
1570 ValTrpMetSerAsnThrGlyTyrGlyArgAspTyrAlaSerAlaGlnTy 1586
3714 CCGCGCCTACCGCCAACAAACCGACCTGCGCCAAATCGGTATGCAGAAAA 3763
1586 rArgArgPheSerSerLysArgThrGlnThrGlnIleGlyIleAspArgS 1603
3764 ACCTCGGCAGCGGG...CGCGTCGGCATCCTGTTTTCGCACAACCGGACC 3810
1603 erLeuSerGluAsnMetGlnIleGlyGlyValLeuThrTyrSerAspSer 1619
3811 GGAAACACCTTCGACGACGGCATCGGCAACTCGGCACGGCTTGCCCACGG 3860
1620 GlnHisThrPheAspGlnAlaSerGlyLysAsnThr...PheValGlnAl 1635
3861 TGCCGTTTTCGGGCAATACGGCATCGGC...AGGTTCGACATCGGCATCA 3907
1635 aAsnLeuTyrGlyLysTyrTyrLeuAsnAspAlaTrpTyrValAlaGlyA 1652
3908 GCGCGGGCGCGGGTTTTAGTAGCGGCAGCCTTTCAGACGGCATCAGAGGC 3957
1652 spIleGlyAlaGlySerLeuArgSerArgLeuGlnThrGlnGlnLysAla 1668
3958 AAAATCCGCCGCCGCGTGCTGCATTACGGCATTCAGGCAAGATACCGCGC 4007
1669 AsnPheAsnArgAlaSerIleGlnThrGlyLeuThrLeuGlyAsnThrLe 1685
4008 AGGTTTCGGCGGATTCGGCATCGAACCGCACATCGGCGCAACGCGCTATT 4057
1685 uLysIleAsnGlnPheGluIleValProSerAlaGlyIleArgTyrSerA 1702
4058 TCGTCCAAAAAGCGGATTACCGATACGAAAACGTCAATATCGCCACCCCG 4107
1702 rgLeuSerSerAlaAspTyrLysLeuGlyAsnAspSerValLysValSer 1718
4108 GGCCTTGCATTCAACCGCTACCGCGCGGGCATTAAGGCAGATTATTCATT 4157
1719 SerMetSerValLysThrLeuThrAlaGlyLeuAspPheAlaTyrArgPh 1735
4158 CAAACCGGCGCAACACATTTCCATCACGCCTTATTTGAGCCTGTCCTAT. 4206
1735 eLys...ValGlyAsnLeuThrValLysProLeuLeuSerAlaAlaTyrP 1751
4207 ..ACCGATGCCGCTTCCGGCAAAGTCCGAACGCGCGTCAATACCGCCGTA 4254
1751 heAlaAsnTyrGlyLysGlyGlyValAsnValGlyGlyAsnSerPheVal 1767
4255 TTGGCGCAGGATTTCGGCAAAACCCGCAGTGCGGAATGGGGCGTAAACGC 4304
1768 TyrLysAlaAspAsnGlnGlnGlnTyrSerAla.....GlyAlaAlaLe 1782
4305 CGAAATCAAAGGTTTCACGCTGTCCCTCCACGCTGCCGCCGCCAAGGGGC 4354
1782 uLeuTyrArgAsnValThrLeuAsnValAsnGlySerIleThrLysGlyL 1799
4355 CGCAATTGGAAGCGCAGCACAGCGCGGGCATCAAATTAGGCTACCGCTGG 4404
1799 ysGlnLeuGluLysGlnLysSerGlyGlnIleLysIleGlnIleArgPhe 1815

```
seq_name: pir2:I54632

seq_documentation_block:
tsh protein - Escherichia coli
C;Species: Escherichia coli
C;Date: 07-Jun-1996 #sequence_revision 07-Jun-1996 #text_change 08-Dec-2000
C;Accession: I54632
R;Provence, D.L.; Curtiss, R.
Infect. Immun. 62, 1369-1380, 1994
A;Title: Isolation and characterization of a gene involved in hemagglutination by an avi
A;Reference number: I54632; MUID:94178945
A;Accession: I54632
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-1377 <RES>
A;Cross-references: GB:L27423; NID:g469235; PIDN:AAA24698.1; PID:g469236
C;Superfamily: IgA-specific metalloendopeptidase


alignment_scores:
           Quality: 1125.50            Length:   1529
             Ratio:  1.373              Gaps:     65
 Percent Similarity:  53.630   Percent Identity: 26.226

alignment_block:
 US-09-303-518D-653 x I54632   ..

 Align seq 1/1  to: I54632  from: 1  to: 1377

      82 TTAGCCATATGCCTGTCGTTCGGCATTCTGCCCCAAGCCCGGGCGGGACA 131
         :::   :::   :::|||   |||   |||            ||||||
      42 IleProValLeuPheSerAlaGlySerLeu............AlaGlyTh 54

     132 CACTTATTTCGGCATCAACTACCAATACTATCGCGACTTTGCCGAAAATA 181
         |:::         ::::::||||||   :::||||||||||||||||||
      54 rValAsnAsnGluLeuGlyTyrGlnLeuPheArgAspPheAlaGluAsnL 71

     182 AAGGCAAGTTTGCAGTCGGGGCGAAAGATATTGAGGTTTACAACAAAAAA 231
         |||||   |||      ||||||   :::|||   :::|||||||||:::
      71 ysGlyMetPheArgProGlyAlaThrAsnIleAlaIleTyrAsnLysGln 87

     232 CGGGAGTTGGTCGGCAAATCGATGACGAAAGCCCCGATGATTGATTTTTC 281
         ||||||:::||||||   ::::::   ||||||   |||   ||||||||
      88 GlyGluPheValGly...ThrLeuAspLysAlaAlaMetProAspPheSe 103

     282 TGTGGTA...TCGCGTAACGGCGTGGCGGCATTGGCGGGCGATCAATATA 328
         |:::|||   |||:::   ||||||||||:::|||   :::   |||||||
     103 rAlaValAspSerGluIleGlyValAlaThrLeuIleAsnProGlnTyrI 120

     329 TTGTGAGCGTGGCACATAACGGCGGCTATAACAATGTTGATTTTGGTGCG 378
         ||:::||||||   ||||||||||||||:::|||||||:::||||||
     120 leAlaSerValLysHisAsnGlyGlyTyrThrAsnValSerPheGly... 135

     379 GAGGGAAGCAATCCCGATCAGCACCGCTTTTCTTACCAAATTGTGAAAAG 428
         :::|||:::|||                     |||:::||||||   ||
     136 AspGlyGluAsn.................ArgTyrAsnIleValAspAr 146

     429 AAATAATTATAAAGCAGGGACTAACGGCCATCCTTATGGCGGCGATTATC 478
         |||||||      :::                              |||:::|
     146 gAsnAsnAlaProSer.....................LeuAspPheH 155

     479 ATATGCCGCGTTTGCACAAATTTGTCACAGATGCAGAACCTGTTGAGATG 528
         ||   |||||||||   |||:::||||||:::::   |||:::   :::
     155 isAlaProArgLeuAspLysLeuValThrGluValAlaProThrAlaVal 171

     529 ACCAGTTAT...ATGGATGGGTGGAAATACGCTGATTTAAATAAATACCC 575
         |||:::            :::      |||   |||   ::::::|||||
     172 ThrAlaGlnGlyAlaValAlaGlyAlaTyrLeuAspLysGluArgTyrPr 188

     576 CGATCGTGTTCGAATCGGAGCAGGCAGACAATATTGGCGGTCTGATGAAG 625
                 |||:::|||:::|||   ||||||   ::::::::::::
     188 oValPheTyrArgLeuGlySerGlyThrGlnTyrIleLysAspSerAsnG 205

     626 ACGAACCCAATAACCGCGAAAGTTCATATCATATTGCAAGCGCATATTCT 675
           :::   :::                    :::::::::::|||||||||
     205 lyGlnLeuThr.................GlnMetGlyGlyAlaTyrSer 215

     676 TGGCTCGTCGGTGGCAATACCTTTGCA............CAAAATGGA.. 711
         ||||||:::||||||::::::   :::           |||||||||
     216 TrpLeuThrGlyGlyThrValGlySerLeuSerSerTyrGlnAsnGlyGl 232

     712 .........TCAGGTGGTGGCACAGTCAACTTAGGTAGCGAAAAAATTA 751
                 :::::::::|||   |||                   :::
     232 uMetIleSerThrSerSerGlyLeuVal.................PheA 243

     752 AACATAGCCCATATGGTTTTTTACCAACAGGAGGCTCATTTGGCGACAGT 801
           ::::::      |||   :::|||      |||:::   |||||||||
     243 spTyrLysLeuAsnGlyAlaMetProIleTyrGlyGluAlaGlyAspSer 259

     802 GGCTCACCAATGTTTATCTATGATGCCCAAAAGCAAAAGTGGTTAATTAA 851
         |||||||||:::|||   :::|||:::   ::::::||||||::::::
     260 GlySerProLeuPheAlaPheAspThrValGlnAsnLysTrpValLeuVa 276

     852 TGGGGTATTGCAAACAGGCAACCCCTATATAGGAAAAAGCAATGGCTTCC 901
         |||||||||   :::|||||||         |||:::::|||:::::
     276 lGlyValLeuThrAlaGlyAsnGlyAlaGlyGlyArgGlyAsnAsnTrpA 293

     902 AGCTAGTTCGTAAAGATTGGTTCTATGATGAAATCTTTGCTGGAGATACC 951
           ::::::      |||::::::  :::::::::          :::
     293 laValIleProLeuAspPheIleGlyGlnLysPheAsnGluAspAsnAsp 309

     952 CATTCAGTATTCTACGAACCACATCAAAATGGG......AAATACTTTTT 995
               |||   ::::::      :::::::|||      ::::::   ||
     310 AlaProValThrPheArgThrSerGluGlyGlyAlaLeuGluTrpSerPh 326

     996 TAACGACAATAATAATGGCGCAGGAAAAATCGATGCCAAACATAAACACT 1045
         ||||   :::::::::|||||||||   :::                 |
     326 eAsn...SerSerThrGlyAlaGlyAlaLeuThrGlnGlyThrThrThrT 342

    1046 ATTCTCTACCTTATAGATTAAAAACACGAACCGTTCAATTGTTTAATGTT 1095
         ||::::::
     342 yrAlaMet..................................... 344

    1096 TCTTTATCCGAGACAGCAAGAGAACCTGTTTATCATGCTGCAGGTGGGGT 1145
                                          |||::      |||
     345 ..............................HisGlyGlnGlnGlyAs 350

    1146 CAACAGTTATCGACCCAGACTGAATAATGGAGAAAATATTTCCTTTATTG 1195
           :::                 ||||||   |||:::|||:::   |||
     350 nAsp.............LeuAsnAlaGlyLysAsnLeuIlePheGlnG 362

    1196 ACAAAGGAAAAGGTGAATTGATACTTACCAGCAACATCAACCAAGGCGCG 1245
           :::   :::|||::::::   |||   ::::::::::::|||||||||
     362 lyGln...AsnGlyGlnIleAsnLeuLysAspSerValSerGlnGlyAla 377

    1246 GGCGGTTTGTATTTTGAGGGTAATTTTACGGTCTCGCCTAAAAACAACGA 1295
         |||:::|||   |||:::   |||:::||||||:::   :::|||:::::
     378 GlySerLeuThrPheArgAspAsnTyrThrValThrThrSerAsnGlySe 394

    1296 AACGTGGCAAGGCGCGGGCGTTCATATCAGTGATGGCAGTACCGTTACTT 1345
         :||||||   ||||||||||:::   :::::::::|||   :::|||:::|
     394 rThrTrpThrGlyAlaGlyIleValValAspAsnGlyValSerValAsnT 411

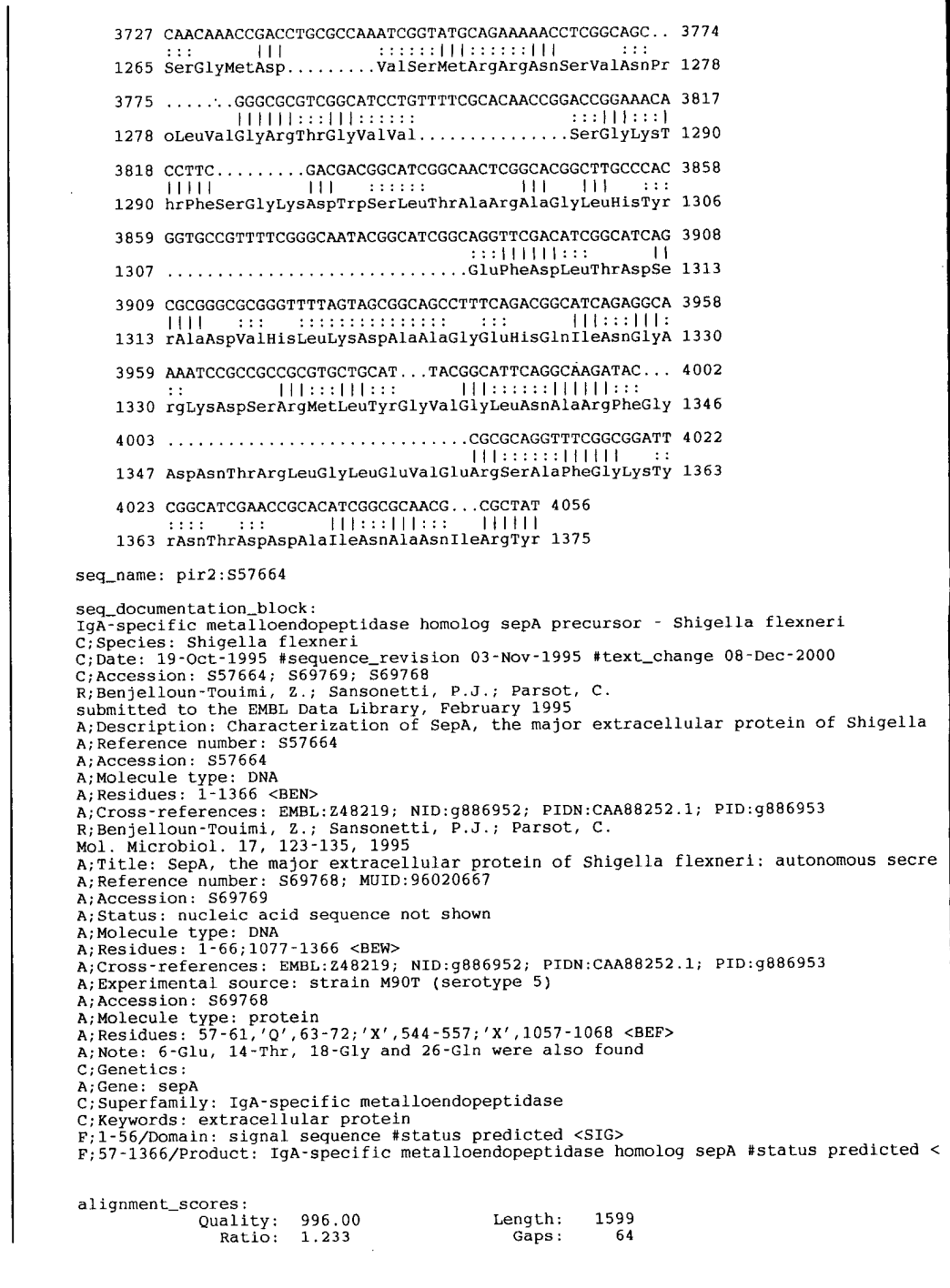
    1346 GGAAAGTAAACGGCGTGGCAAACGACCGCCTGTCCAAAATCGGCAAAGGC 1395
         ||:::|||||||||||||   :::|||:::|||   |||||||||:::|||
     411 rpGlnValAsnGlyValLysGlyAspAsnLeuHisLysIleGlyGluGly 427

    1396 ACGCTGCTGGTTCAAGCCAAAGGGGAAAACCAAGGCTCGGTCAGCGTGGG 1445
         ||||||   |||||||:::   |||   |||:::|||:::::::::|||||
     428 ThrLeuThrValGlnGlyThrGlyIleAsnGluGlyGlyLeuLysValGl 444

    1446 CGACGGTAAAGTCATCTTAGATCAGCAGGCGGACGATCAAGGCAAAAAAC 1495
         ||||||||||||||:::|||:::||||||||||||||:::::|||:::   |
     444 yAspGlyLysValValLeuAsnGlnGlnAlaAspAsnLysGlyGlnValG 461
```







seq_name: pir2:S57664

seq_documentation_block:
IgA-specific metalloendopeptidase homolog sepA precursor - Shigella flexneri
C;Species: Shigella flexneri
C;Date: 19-Oct-1995 #sequence_revision 03-Nov-1995 #text_change 08-Dec-2000
C;Accession: S57664; S69769; S69768
R;Benjelloun-Touimi, Z.; Sansonetti, P.J.; Parsot, C.
submitted to the EMBL Data Library, February 1995
A;Description: Characterization of SepA, the major extracellular protein of Shigella
A;Reference number: S57664
A;Accession: S57664
A;Molecule type: DNA
A;Residues: 1-1366 <BEN>
A;Cross-references: EMBL:Z48219; NID:g886952; PIDN:CAA88252.1; PID:g886953
R;Benjelloun-Touimi, Z.; Sansonetti, P.J.; Parsot, C.
Mol. Microbiol. 17, 123-135, 1995
A;Title: SepA, the major extracellular protein of Shigella flexneri: autonomous secre
A;Reference number: S69768; MUID:96020667
A;Accession: S69769
A;Status: nucleic acid sequence not shown
A;Molecule type: DNA
A;Residues: 1-66;1077-1366 <BEW>
A;Cross-references: EMBL:Z48219; NID:g886952; PIDN:CAA88252.1; PID:g886953
A;Experimental source: strain M90T (serotype 5)
A;Accession: S69768
A;Molecule type: protein
A;Residues: 57-61,'Q',63-72;'X',544-557;'X',1057-1068 <BEF>
A;Note: 6-Glu, 14-Thr, 18-Gly and 26-Gln were also found
C;Genetics:
A;Gene: sepA
C;Superfamily: IgA-specific metalloendopeptidase
C;Keywords: extracellular protein
F;1-56/Domain: signal sequence #status predicted <SIG>
F;57-1366/Product: IgA-specific metalloendopeptidase homolog sepA #status predicted <

alignment_scores:
Quality: 996.00 Length: 1599
Ratio: 1.233 Gaps: 64

Percent Similarity: 50.532 Percent Identity: 25.203

alignment_block:
US-09-303-518D-653 x S57664 ..

Align seg 1/1 to: S57664 from: 1 to: 1366

```
  22 ACAACCGAAACACACCGCAAAGCCCCTAAAACCGGCCGCATCCGCTTCTC 71
     |||   ::::::||||||:::      :::         |||      ||
  27 ThrCysLysSerHisArgArgLeuSerArgArg...ValIleLeuThrSe 42

  72 GCCCGCTTACTTAGCCATATGCCTGTCGTTCGGCATTCTGCCCCAAGCCC 121
     |   |||   |||::::::      ::::::      |||       :::
  42 rValAlaAlaLeuSerLeuSerSerAlaTrpProAlaLeuSerAlaThrV 59

 122 GGGCGGGACACACTTATTTCGGCATCAACTACCAATACTATCGCGACTTT 171
       :::::::::            |||   ||||||   :::||||||||||
  59 alSerAlaGlu............IleProTyrGlnIlePheArgAspPhe 71

 172 GCCGAAAATAAAGGCAAGTTTGCAGTCGGGGCGAAAGATATTGAGGTTTA 221
     ||||||||||||||:::|||:::   |||:::   :::|||:::::::||
  72 AlaGluAsnLysGlyGlnPheThrProGlyThrThrAsnIleSerIleTy 88

 222 CAACAAAAAAGGGGAGTTGGTCGGCAAATCGATGACGAAAGCCCCGATGA 271
     |:::|||:::|||:::||||||||||||   :::   ||||||||||||
  88 rAspLysGlnGlyAsnLeuValGlyLys...LeuAspLysAlaProMetA 104

 272 TTGATTTTTCTGTGGTATCGCGTAACGGCGTGGCGGCATTGGCGGGCGAT 321
       |||||||||   ::::::   :::      :::         ||||||
 104 laAspPheSerSerAlaThrIleThrThrGlySerLeuProProGlyAsp 120

 322 .................CAATATATTGTGAGCGTGGCACATAACGGCGG 353
                      ||||||:::|||:::::   |||   :::||
 121 HisThrLeuTyrSerProGlnTyrValValThrAlaLysHisValSerGl 137

 354 CTATAACAATGTTGATTTTGGTGCGGAGGGAAGCAATCCCGATCAGCACC 403
     |   ::::::::::::||||||
 137 ySerAspThrMetSerPheGly..................... 144

 404 GCTTTTCTTACCAAATTGTGAAAAGAAATAATTATAAAGCA...GGGACT 450
                          :::|||:::|||   |||   ||||||
 145 .................TyrAlaLysAsnThrTyrThrAlaValGlyThr 155

 451 AACGGCCATCCTTATGGCGGCGATTATCATATGCCGCGTTTGCACAAATT 500
     |||::::::      |||   |||            ||||||   |||::
 156 AsnAsnAsnSer...GlyLeuAspIleLysThrArgArgLeuSerLysLe 171

 501 TGTCACAGATGCAGAACCTGTTGAGATGACC.................. 531
     :||||||::::::   |||:::|||::::::
 171 uValThrGluValAlaProAlaGluValSerAspIleGlyAlaValSerG 188

 532 ..AGTTATATGGATGGGTGGAAATACGCTGATTTAAATAAATACCCTGAT 579
       :::|||:::   |||   :::::::::::::
 188 lyAlaTyrGlnAlaGlyGlyArgPheThrGluPhe.............. 199

 580 CGTGTTCGAATCGGAGCAGGCAGACAATATTGGCGGTCTGATGAAGACGA 629
           |||:::|||:::|||   ||||||   :::         |||::
 200 ...TyrArgLeuGlyGlyGlyMetGlnTyrValLys.........AspLy 212

 630 ACCCAATAACCGCGAAAGTTCATATCATATTGCAAGCGCATATTCTTGGC 679
     :   :::||||||   :::   |||         :::      ::::::|
 212 sAsnGlyAsnArgThrGlnValTyr.........ThrAsnGlyGlyPheL 226

 680 TCGTCGGTGGCAATACCTTTGCACAAAATGGATCAGGTGGTGGC...... 723
     ||||||||||:::::   |||   |||:::   :::::::|||
 226 euValGlyGlyThrValSerAlaLeuAsnSerTyrAsnAsnGlyGlnMet 242

 724 ...ACAGTCAACTTAGGTAGCGAAAAAATTAAACATAGCCCATATGGTTT 770
        |||::::::   |||   :::   :::         |||   |||
 243 IleThrAlaGlnThrGly...AspIlePheLeuIleProProAsnGlyPr 258

 771 TTTACCAACAGGAGGCTCATTTGGCGACAGTGGCTCACCAATGTTTATCT 820
      |||   :::      ::::::|||||||||||||||||||:::|||   |
```

```
 258 oLeuAlaAsnTyrLeuAsnMetGlyAspSerGlySerProLeuPheAlaT 275

 821 ATGATGCCCAAAAGCAAAAGTGGTTAATTAATGGGGTATTGCAAACAGGC 870
     |||||:::   ::::::||||||:::::::   |||||||||:::::|||
 275 yrAspSerLeuGlnLysLysTrpValLeuIleGlyValLeuSerSerGly 291

 871 AACCCCTATATAGGAAAAAGCAATGGCTTCCAGCTAGTTCGTAAAGATTG 920
     :::   |||         :::|||::::::   ::::::   :::|||::
 292 ThrAsnTyr.........GlyAsnAsnTrpValValThrThrGlnAspPh 305

 921 GTTCTATGATGAAATCTTTGCTGGAGATACCCATTCAGTATTCTACGAAC 970
     ::::
 305 eLeu..................................................... 306

 971 CACATCAAAATGGGAAATACTTTTTTAACGACAATAATAATGGCGCAGGA 1020
               |||:::         ||||||
 307 ...........GlyGlnGlnProGlnAsnAsp..................... 313

1021 AAAATCGATGCCAAACATAAACACTATTCTCTACCTTATAGATTAAAAAC 1070

 313 ..................................................... 313

1071 ACGAACCGTTCAATTGTTTAATGTTTCTTTATCCGAGACAGCAAGAGAAC 1120
                    |||:::   :::::::::   |||:::   |||
 314 ...............PheAspLysThrIleAlaTyrThrSerGlyGluG 325

1121 CTGTT............TATCATGCTGCAGGTGGGGTCAAC......... 1149
       |||            |||   ||||||:::|||::::::
 325 lyValLeuGlnTrpLysTyrAspAlaAlaAsnGlyThrGlyThrLeuThr 341

1150 ........................AGTTATCGACCCAGA.........CT 1166
                             :::::::::   |||         ||
 342 GlnGlyAsnThrThrTrpGlyTyrAlaTrpLysGluArgLysLeuIleLe 358

1167 GAATAATGGAGAAAATATTTCCTTTATTGACAAAGGAAAAGGTGAATTGA 1216
     ||||   |||:::|||:::   |||         ::::::||||||::::
 358 uAsnAlaGlyLysAsnLeuLeuPheThrGly...AsnAsnGlyGluValV 374

1217 TACTTACCAGCAACATCAACCAAGGCGCGGGCGGTTTGTATTTTGAGGGT 1266
     ::|||   :::::::::||||||||||||||||   |||   |||   |||
 374 alLeuGlnAsnSerValAsnGlnGlyAlaGlyTyrLeuGlnPheAlaGly 390

1267 AATTTTACGGTCTCGCCTAAAAACAACGAAACGTGGCAAGGCGCGGGCGT 1316
     ::::::   ||||||      |||::::::||||||:::|||:::|||::
 391 AspTyrArgValSerAlaLeuAsnGlyGlnThrTrpMetGlyGlyGlyIl 407

1317 TCATATCAGTGATGGCAGTACCGTTACTTGGAAAGTAAACGGCGTGGCAA 1366
     :      :::   |||:::   |||   |||:::||||||||||||||||:
 407 eIleThrAspLysGlyThrHisValLeuTrpGlnValAsnGlyValAlaG 424

1367 ACGACCGCCTGTCCAAAATCGGCAAAGGCACGCTGCTGGTTCAAGCCAAA 1416
     ::|||:::|||   |||   |||:::|||||||||   |||::::::
 424 lyAspAsnLeuHisLysThrGlyGluGlyThrLeuThrValAsnGlyThr 440

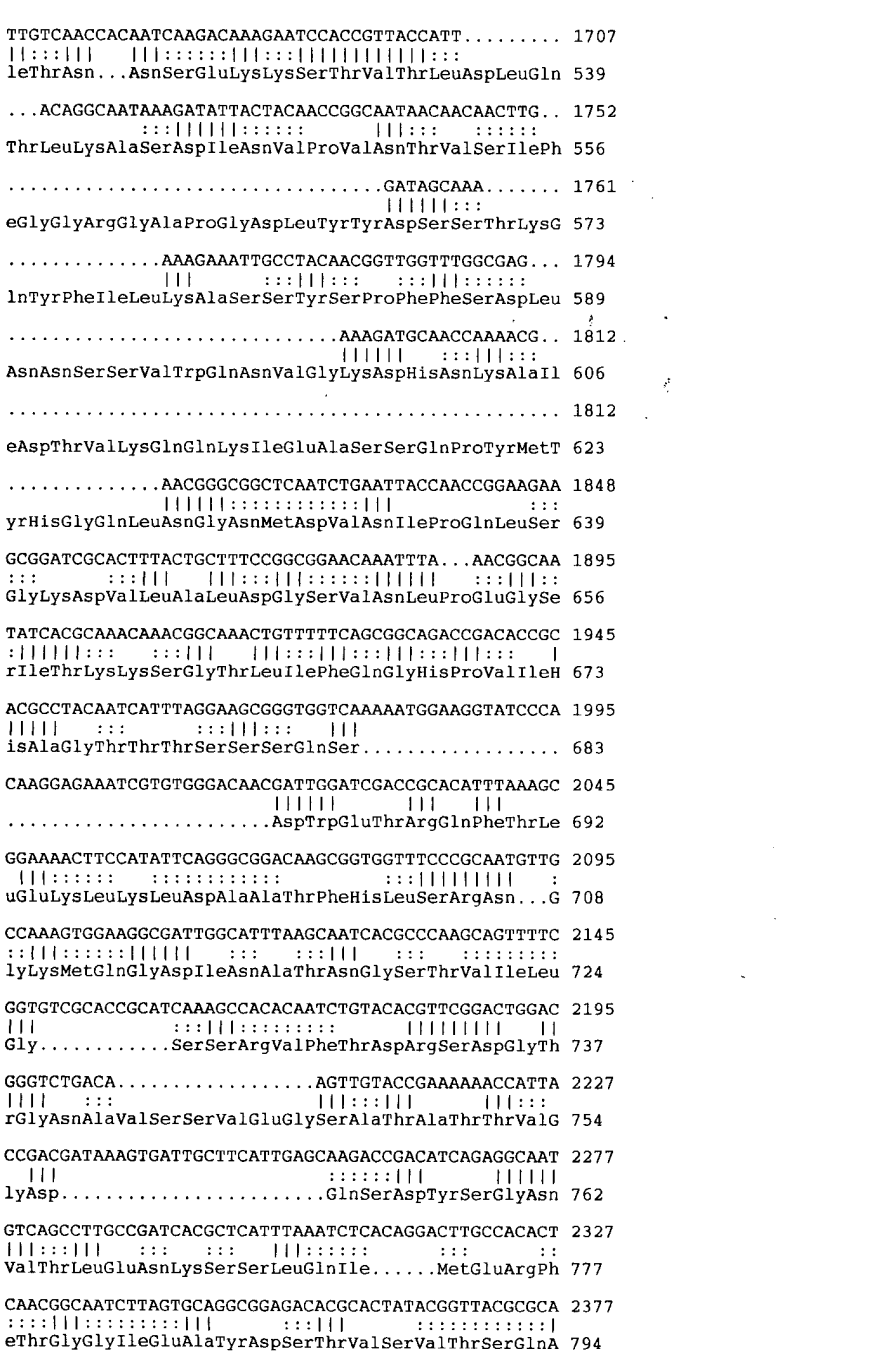
1417 GGGGAAAACCAAGGCTCGGTCAGCGTGGGCGACGGTAAAGTCATCTTAGA 1466
     |||   |||   |||:::::::::||||||||||   |||||||||::
 441 GlyValAsnAlaGlyGlyLeuLysValGlyAspGlyThrValIleLeuAs 457

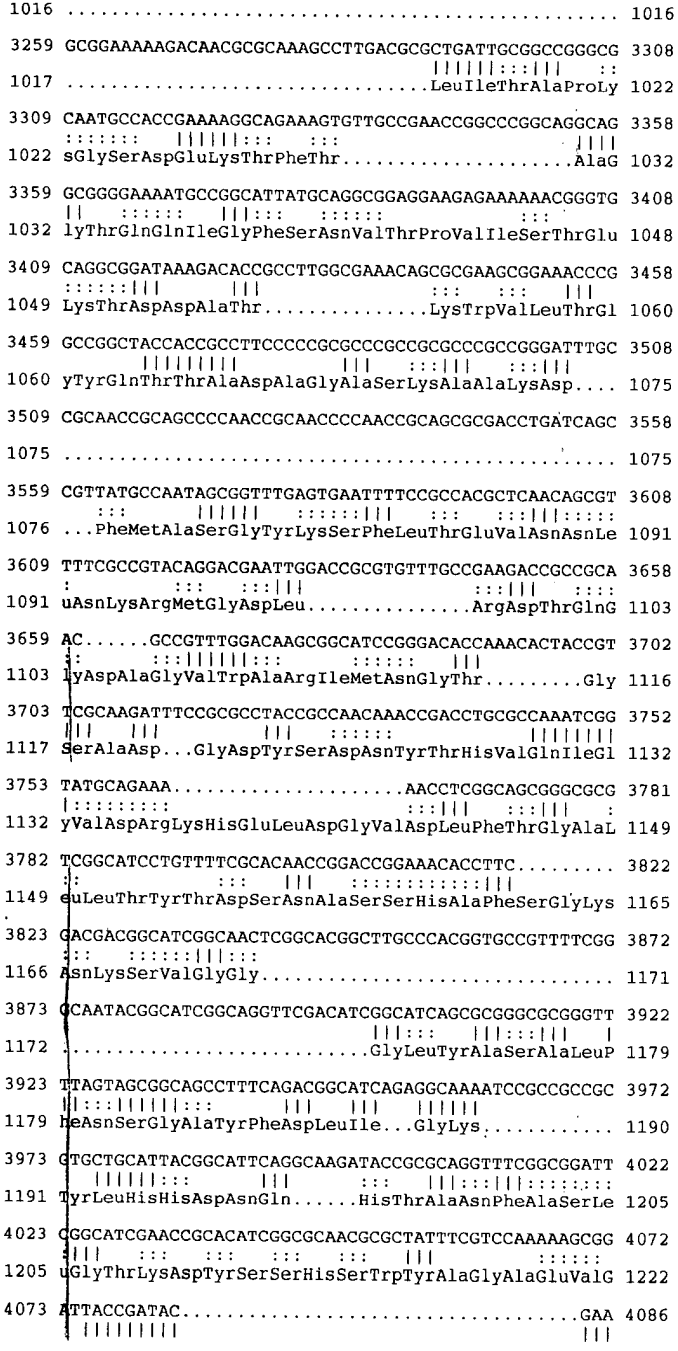
1467 TCAGCAGGCGGACGATCAAGGCAAAAAACAAGCCTTTAGTGAAATCGGCT 1516
     :||||||||||||   :::||||||   |||||||||||:::::::|||:
 457 nGlnGlnAlaAspArgAspGlyLysValGlnAlaPheSerSerValGlyI 474

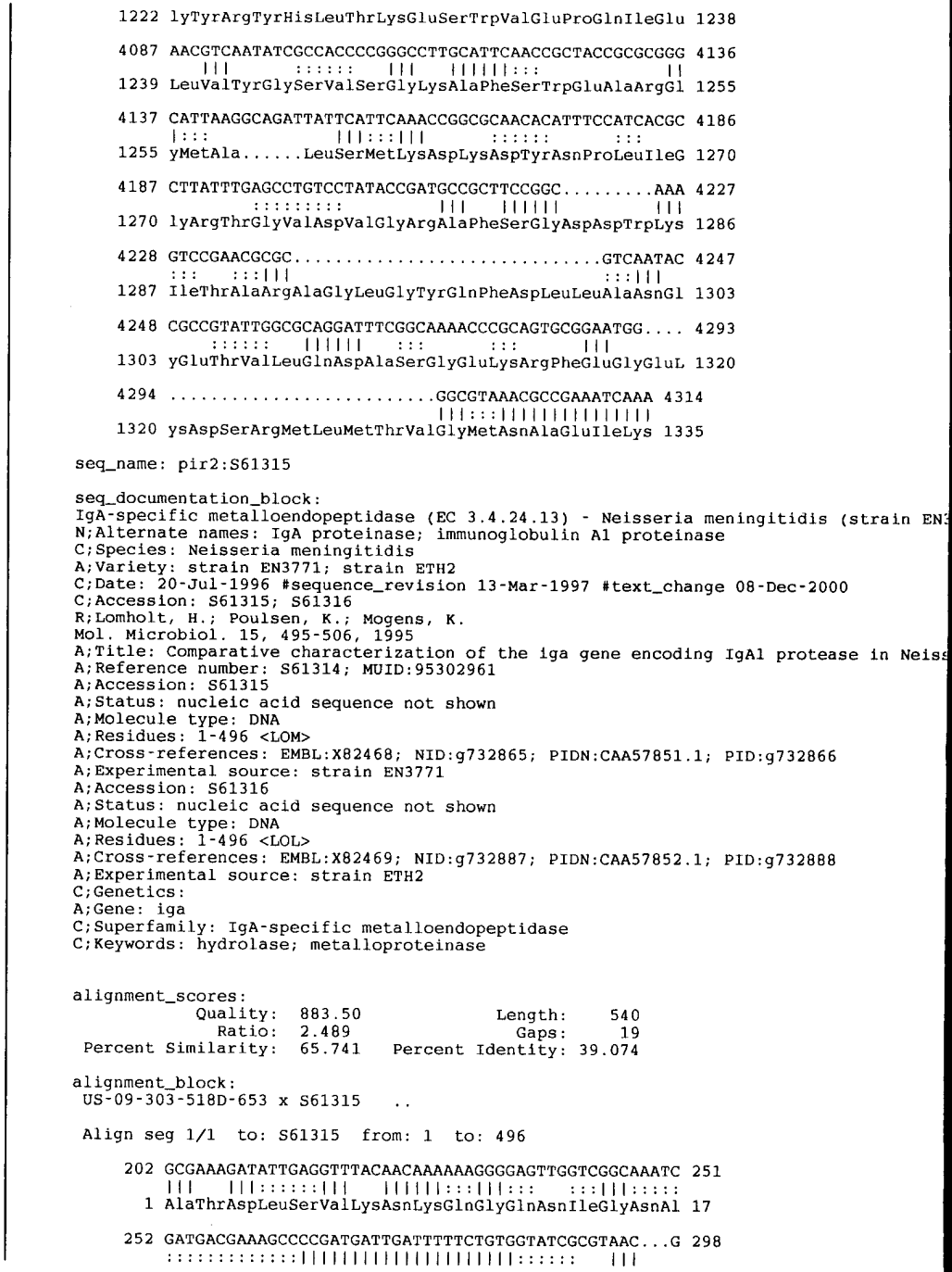
1517 TGGTCAGCGGCAGGGGGACGGTGCAACTGAATGCCGATAATCAGTTCAAC 1566
     :::::|||||||||   ||||||   |||:::   :::::|||   |||
 474 leAlaSerGlyArgProThrValValLeuSerAspSerGlnGlnValAsn 490

1567 CCCGACAAACTCTATTTCGGCTTTCGCGGCGGACGTTTGGATTTGAACGG 1616
     ||||||::::::   :::|||:::||||||||||||||:::|||||||||
 491 ProAspAsnIleSerTrpGlyTyrArgGlyGlyArgLeuGluLeuAsnGl 507

1617 GCATTCGCTTTCGTTCCACCGCATTCAAAATACCGATGAAGGGGCGATGA 1666
     |::::::|||:::|||   |||:::|||   :::|||   |||||||:::|
 507 yAsnAsnLeuThrPheThrArgLeuGlnAlaAlaAspTyrGlyAlaIleI 524
```







```
seq_name: pir2:S61315

seq_documentation_block:
IgA-specific metalloendopeptidase (EC 3.4.24.13) - Neisseria meningitidis (strain EN3
N;Alternate names: IgA proteinase; immunoglobulin A1 proteinase
C;Species: Neisseria meningitidis
A;Variety: strain EN3771; strain ETH2
C;Date: 20-Jul-1996 #sequence_revision 13-Mar-1997 #text_change 08-Dec-2000
C;Accession: S61315; S61316
R;Lomholt, H.; Poulsen, K.; Mogens, K.
Mol. Microbiol. 15, 495-506, 1995
A;Title: Comparative characterization of the iga gene encoding IgA1 protease in Neiss
A;Reference number: S61314; MUID:95302961
A;Accession: S61315
A;Status: nucleic acid sequence not shown
A;Molecule type: DNA
A;Residues: 1-496 <LOM>
A;Cross-references: EMBL:X82468; NID:g732865; PIDN:CAA57851.1; PID:g732866
A;Experimental source: strain EN3771
A;Accession: S61316
A;Status: nucleic acid sequence not shown
A;Molecule type: DNA
A;Residues: 1-496 <LOL>
A;Cross-references: EMBL:X82469; NID:g732887; PIDN:CAA57852.1; PID:g732888
A;Experimental source: strain ETH2
C;Genetics:
A;Gene: iga
C;Superfamily: IgA-specific metalloendopeptidase
C;Keywords: hydrolase; metalloproteinase


alignment_scores:
            Quality:  883.50            Length:    540
              Ratio:  2.489               Gaps:     19
 Percent Similarity:  65.741   Percent Identity: 39.074

alignment_block:
 US-09-303-518D-653 x S61315   ..

 Align seg 1/1  to: S61315  from: 1  to: 496
```



seq_name: pir2:S61319

seq_documentation_block:
IgA-specific metalloendopeptidase (EC 3.4.24.13) homolog - Neisseria meningitidis (fr
C;Species: Neisseria meningitidis
A;Variety: NG117
C;Date: 20-Jul-1996 #sequence_revision 13-Mar-1997 #text_change 08-Dec-2000
C;Accession: S61319
R;Lomholt, H.; Poulsen, K.; Mogens, K.
Mol. Microbiol. 15, 495-506, 1995
A;Title: Comparative characterization of the iga gene encoding IgA1 protease in Neiss
A;Reference number: S61314; MUID:95302961
A;Accession: S61319
A;Status: preliminary; nucleic acid sequence not shown
A;Molecule type: DNA
A;Residues: 1-508 <LOM>
A;Cross-references: EMBL:X82483; NID:g732881; PIDN:CAA57866.1; PID:g732882
C;Superfamily: IgA-specific metalloendopeptidase
C;Keywords: hydrolase; metalloproteinase

```
alignment_scores:
           Quality:  876.50              Length:    558
             Ratio:  2.483                 Gaps:     17
 Percent Similarity:  63.262   Percent Identity: 37.634

alignment_block:
 US-09-303-518D-653 x S61319   ..

 Align seg 1/1  to: S61319  from: 1  to: 508

     202 GCGAAAGATATTGAGGTTTACAACAAAAAAGGGGAGTTGGTCGGCAAATC 251
         |||   |||::::::|||   ||||||:::|||:::   :::|||:::::
       1 AlaThrAspLeuSerValLysAsnLysGlnGlyGlnAsnIleGlyAsnAl 17

     252 GATGACGAAAGCCCCGATGATTGATTTTTCTGTGGTATCGCGTAAC...G 298
         :::::::::::::|||||||||||||||||||||::::::   |||
      17 aLeuSerAsnValProMetIleAspPheSerValAlaAspValAsnLysA 34

     299 GCGTGGCGGCATTGGCGGGCGATCAATATATTGTGAGCGTGGCACAT... 345
           :::|||:::::::::      ||||||   |||||||||   |||.
      34 rgIleAlaThrValValAspProGlnTyrAlaValSerValLysHisAla 50

     346 .......................AACGGCGGCTATAACAATGTTGATTT 371
                                   |||   ||||||:::   :::
      51 LysAlaGluValHisThrPheTyrTyrGlyGlnTyrAsnGlyHisAsnAs 67

     372 TGGTGCGGAGGGAAGCAATCCCGATCAGCACCGCTTTTCTTACCAAATTG 421
             |||:::   :::|||                 :::|||::::::|
      67 pValAlaAspLysGluAsn.................GluTyrArgValV 78

     422 TGAAAAGAAATAATTATAAAGCAGGGACTAACGGCCATCCTTATGGCGGC 471
         ||::::::||||||||||:::                |||         |||
      78 alGluGlnAsnAsnTyrGluPro............HisLysAlaTrpGly 90

     472 .......................GATTATCATATGCCGCGTTTGCACAA 497
                                 ||||||:::|||   |||::::::||
      91 AlaSerAsnLeuGlyArgLeuGluAspTyrAsnMetAlaArgPheAsnLy 107

     498 ATTTGTCACAGATGCAGAACCTGTTGAGATGACCAGTTATATGGATGGGT 547
         ||||||||||::::::   |||:::      |||:::         |||
     107 sPheValThrGluValAlaProIleAlaProThrAspAlaGlyGlyGlyL 124

     548 GG...AAATACGCTGATTTAAATAAATACCCTGATCGTGTTCGAATCGGA 594
              |||   |||   |||::::::   :::   ||||||:::|||
     124 euAspThrTyrLysAspLysAsnArgPheSerSerPheValArgValGly 140

     595 GCAGGCAGACAA......TATTGGCGGTCTGATGAAGACGAACCCAATAA 638
         ||||||||||||      |||   ::::::         :::   :::||
     141 AlaGlyArgGlnLeuValTyrGluLysGlyAlaTyrHisGlnGluGlyAs 157

     639 CCGCGAAAGTTCATAT.........CATATTGCAAGCGCATATTCTTGGC 679
         |   |||::::::|||           ::::::::::||||||   :::
     157 n...GluLysGlyTyrAspLeuArgAspLeuSerGlnAlaTyrArgTyrA 173

     680 TCGTCGGTGGC.................................AAT 693
           ::::::|||                                  |||
     173 laIleAlaGlyThrProTyrLysAspIleAsnIleAspGlnThrMetAsn 189

     694 ACCTTTGCACAAAATGGATCAGGTGGTGGCACAGTCAACTTAGGTAGCGA 743
         |||   :::      |||   |||:::   :::::::::   ::::::||
     190 ThrGluGlyLeuIleGlyPheGlyAsnHisAsnThrHisTyrSerAlaGl 206

     744 AAAAATTAAACATAGC...CCATATGGTTTTTTACCAACAGGAGGCTCAT 790
         |:::::::|::::::               |||   :::   |||   :
     206 uGluLeuLysGlnAlaLeuSerGlnAspAlaLeuThrAsnTyrGlyValL 223

     791 TTGGCGACAGTGGCTCACCAATGTTTATCTATGATGCCCAAAAGCAAAAG 840
         ::|||||||||||||||||:::|||   :::|||   ||||||::::::
     223 euGlyAspSerGlySerProLeuPheAlaPheAspLysGlnLysAsnGln 239

     841 TGGTTAATTAATGGGGTATTGCAAACAGGCAACCCCTATATAGGAAAAAG 890
         |||::::::   |||:::   :::            |||      |||||
     240 TrpValPheLeuGlyThrTyrAspTyrTrpAlaGlyTyrGlyLysLysSe 256

     891 CAATGGCTTCCAGCTAGTTCGTAAAGATTGGTTCTATGATGAAATCTTTG 940
         |               :::   |||      |||   |||:::|||
     256 rTrpGlnGluTrpAsnIleTyrLysLysGluPheAlaAspLysIle.... 271

     941 CTGGAGATACCCATTCAGTATTCTACGAACCACATCAAAATGGGAAATAC 990

     271 ..................................................  271

     991 TTTTTTAACGACAATAATAATGGCGCAGGAAAAATCGATGCCAAACATAA 1040
                     :::::::::::::||||||   |||   ::::::::::
     272 .........LysGlnArgAspAsnAlaGlyThrIleLysGlyAsnArgGl 285

    1041 ACACTATTCTCTACCTTATAGATTAAAAACACGAACCGTTCAATTGTTTA 1090
         :|||:::                                       :::|
     285 uHisHis.....................................TrpA 289

    1091 ATGTTTCTTTATCCGAGACAGCAAGAGAACCTGTTTATCATGCTGCAGGT 1140
         ||::::::::::::   :::::::::
     289 snIleThrPheGlyThrAsnSerLys..................... 297

    1141 GGGGTCAACAGTTATCGACCCAGACTG.....................AA 1169
             ::::::|||          ||||||                   ||
     298 ...IleGlySerThrAlaValArgLeuAlaAsnAsnGluArgAspAlaAs 313

    1170 TAATGGAGAAAATATTTCCTTTATTGACAAAGGAAAAGGTGAATTGATAC 1219
         |||||||:::|||::::::|||   |||       :::|||   |||:::|
     313 nAsnGlyGlnAsnValThrPheGluAsp......AsnGlyThrLeuValL 328

    1220 TTACCAGCAACATCAACCAAGGCGCGGGCGGTTTGTATTTTGAGGGTAAT 1269
         ||   :::|||||||||||||||||||||||||||:::|||:::|||:::
     328 euAspGlnAsnIleAsnGlnGlyAlaGlyGlyLeuPhePheLysGlyAsp 344

    1270 TTTACGGTCTCGCCTAAAAACAACGAA...ACGTGGCAAGGCGCGGGCGT 1316
         :::||||||:::      ||||||:::   ||||||   |||||||||::
     345 TyrThrValLysGlyAlaAsnAsnAspIleThrTrpLeuGlyAlaGlyIl 361

    1317 TCATATCAGTGATGGCAGTACCGTTACTTGGAAAGTAAACGGCGTGGCAA 1366
         :   ::::::||||||:::   |||:::|||:::|||::::::      :
     361 eAspValAlaAspGlyLysLysValValTrpGlnValLysAsnProAsnG 378

    1367 ACGACCGCCTGTCCAAAATCGGCAAAGGCACGCTGCTGGTTCAAGCCAAA 1416
         ::|||||||||:::|||||||||||||||||||   :::::::::
     378 lyAspArgLeuAlaLysIleGlyLysGlyThrLeuGluIleAsnGlyThr 394

    1417 GGGGAAAACCAAGGCTCGGTCAGCGTGGGCGACGGTAAAGTCATCTTAGA 1466
         |||   |||||||||:::::::::|||||||||||   |||||||||::
     395 GlyValAsnGlnGlyGlnLeuLysValGlyAspGlyThrValIleLeuAs 411

    1467 TCAGCAGGCGGACGATCAAGGCAAAAAACAAGCCTTTAGTGAAATCGGCT 1516
         :||||||||||||   :::   |||   ||||||||||||:::::::|||:
     411 nGlnGlnAlaAspAlaAspLysLysValGlnAlaPheSerGlnValGlyI 428

    1517 TGGTCAGCGGCAGGGGGACGGTGCAACTGAATGCCGATAATCAGTTCAAC 1566
         ::|||||||||||||||||:::   ||||||::::::||||||:::|||
     428 leValSerGlyArgGlyThrLeuValLeuAsnSerSerAsnGlnIleAsn 444

    1567 CCCGACAAACTCTATTTCGGCTTTCGCGGCGGACGTTTGGATTTGAACGG 1616
         ||||||:::||||||||||||||||||||||||||||||||   |||||
     445 ProAspAsnLeuTyrPheGlyPheArgGlyGlyArgLeuAspAlaAsnGl 461

    1617 GCATTCGCTTTCGTTCCACCGCATTCAAAATACCGATGAAGGGGCGATGA 1666
         |::::::|||:::|||::::::|||:::|||:::||||||||||||   |
     461 yAsnAspLeuThrPheGluHisIleArgAsnValAspGluGlyAlaArgI 478

    1667 TTGTCAACCACAATCAAGACAAAGAATCCACCGTTACCATTACAGGCAAT 1716
         ||||||||||||||   |||:::   ||||||:::|||:::|||||||:::
     478 leValAsnHisAsnThrAspArgAlaSerThrIleThrLeuThrGlyLys 494

    1717 AAAGATATTACTACAACCGGCAAT 1740
         :::   ||||||:::      |||
     495 SerLeuIleThrAlaProGlnAsn 502
```

```
seq_name: pir2:S61328

seq_documentation_block:
IgA-specific metalloendopeptidase (EC 3.4.24.13) homolog - Neisseria meningitidis (fragm
C;Species: Neisseria meningitidis
A;Variety: NG44/76
C;Date: 14-Feb-1997 #sequence_revision 13-Mar-1997 #text_change 08-Dec-2000
C;Accession: S61328
R;Lomholt, H.; Poulsen, K.; Mogens, K.
Mol. Microbiol. 15, 495-506, 1995
A;Title: Comparative characterization of the iga gene encoding IgA1 protease in Neisseri
A;Reference number: S61314; MUID:95302961
A;Accession: S61328
A;Status: preliminary; nucleic acid sequence not shown
A;Molecule type: DNA
A;Residues: 1-497 <LOM>
A;Cross-references: EMBL:X82481
C;Superfamily: IgA-specific metalloendopeptidase
C;Keywords: hydrolase; metalloproteinase


alignment_scores:
            Quality:  875.50              Length:    551
              Ratio:  2.459                Gaps:     20
 Percent Similarity:  64.610   Percent Identity: 38.294

alignment_block:
 US-09-303-518D-653 x S61328   ..

 Align seg 1/1  to: S61328  from: 1  to: 497

     202 GCGAAAGATATTGAGGTTTACAACAAAAAAGGGGAGTTGGTCGGCAAATC 251
         |||   |||:::::|||   ||||||:::|||:::   :::|||:::::
       1 AlaThrAspLeuSerValLysAsnLysGlnGlyGlnAsnIleGlyAsnAl 17

     252 GATGACGAAAGCCCCGATGATTGATTTTTCTGTGGTATCGCGTAAC...G 298
         :::::::::::::|||||||||||||||||||||::::::   |||
      17 aLeuSerAsnValProMetIleAspPheSerValAlaAspValAsnArgA 34

     299 GCGTGGCGGCATTGGCGGGCGATCAATATATTGTGAGCGTGGCACATAAC 348
           :::   ::::::        ||||||   |||||||||   |||
      34 rgThrLeuThrValIleAspProGlnTyrAlaValSerValLysHisVal 50

     349 GGCGGC..............TATAACAATGTTGATTTTGGTGCGGAGGG 383
           |||               |||:::::   :::         :::
      51 LysGlyAspGluIleSerTyrTyrGlyHisHisAsnGlyHisLeuAspVa 67

     384 AAGCAATCCCGATCAGCACCGCTTTTCTTACCAAATTGTGAAAAGAAATA 433
         ||||||   ::::::         :::|||:::   |||   :::|||:
      67 lSerAsnAspGluAsn.........GluTyrArgSerValAlaGlnAsnA 81

     434 ATTATAAAGCAGGGACTAACGGCCATCCTTATGGCGGC............ 471
         ::|||:::   :::   |||   |||   :::|||:::
      81 spTyrGluProAsnLysAsnTrpHis...HisGlyAsnGlnGlyArgLeu 96

     472 ...GATTATCATATGCCGCGTTTGCACAAATTTGTCACAGATGCAGAACC 518
           ||||||:::|||   ||||||:::||||||||||||::::::   ||
      97 GluAspTyrAsnMetAlaArgLeuAsnLysPheValThrGluValAlaPr 113

     519 TGTTGAGATGACCAGTTATATGGATGGGTGGAAA...TACGCTGATTTAA 565
         |:::      ||||||          |||   :::   |||   |||   |
     113 oIleAlaProThrSerAlaGlyGlyGlyValGluThrTyrLysAspLysA 130

     566 ATAAATACCCTGATCGTGTTCGAATCGGAGCAGGCAGACAATATTGGCGG 615
         ||::::::   :::   ||||||:::|||||||||   |||:::
     130 snArgPheSerGluPheValArgValGlyAlaGlyThrGlnPhe...... 144

     616 TCTGATGAAGACGAACCCAATAACCGCGAAAGTTCATATCATATTGCAAG 665
                   |||   |||:::|||   :::      :::::::::
     145 ............GluTyrAsnSerArgTyrAsnMetThrGluLeuSerAr 157

     666 CGCATATTCTTGGCTCGTCGGTGGCAATACCTTT..............G 700
         |||||||::::::|||:::||||||||||||||||||:::   ||||||:

         ||||||   :::   ::::::|||:::   :::               :
     157 gAlaTyrArgTyrAlaIleAlaGlyThrProTyrGlnAspValAsnValT 174

     701 CACAAAATGGATCAGGTGGTGGCACAGTCAACTTAGGTAGCGAAAAAATT 750
         :::::|||   :::       |||   :::::::::|||:::::::::
     174 hrSerAsnLeuAsnGlnGluGlyLeuIleGlyPheGlyAspAsnSerLys 190

     751 AAACATAGCCCA.....................TATGGTTTTTTACC 776
           |||||||||                                :::   |||
     191 HisHisSerProGluLysLeuLysGluValLeuSerGlnAsnAlaLeuTh 207

     777 AACAGGAGGCTCATTTGGCGACAGTGGCTCACCAATGTTTATCTATGATG 826
          :::   :::   :::|||||||||||||||||:::|||   ||||||
     207 rAsnTyrAlaValLeuGlyAspSerGlySerProLeuPheAlaTyrAspL 224

     827 CCCAAAAGCAAAAGTGGTTAATTAATGGGGTATTGCAAACAGGCAACCCC 876
           |||::::::::::|||:::::   |||:::   :::
     224 ysGlnGluLysArgTrpValPheLeuGlyAlaTyrAspTyrTrpAlaGly 240

     877 TATATAGGAAAAAGCAATGGCTTCCAGCTAGTTCGTAAAGATTGG...TT 923
         |||       :::::::|||:::::::|||            :::|||   ::
     241 Tyr......GlnLysAsnSerTrpGln............GluTrpAsnIl 251

     924 CTATGATGAAATCTTTGCTGGAGATACCCATTCAGTATTCTACGAACCAC 973
         :|||   :::   ||||||
     251 eTyrLysLysGluPheAla.............................. 257

     974 ATCAAAATGGGAAATACTTTTTTAACGACAATAATAATGGCGCAGGAAAA 1023
                   |||         :::||||||         ||||||
     258 ........AspLysIleLysGlnArgAspAsn.........AlaGlyThr 268

    1024 ATCGATGCCAAACATAAACACTATTCTCTACCTTATAGATTAAAAACACG 1073
         |||   ::::::   :::|||:::
     269 IleLysGlyAsnGlyGluHisHis........................... 276

    1074 AACCGTTCAATTGTTTAATGTTTCTTTATCCGAGACAGCAAGAGAACCTG 1123
                  :::|||:::::::::::   :::::::::
     277 .............TrpAsnIleThrPheGlyThrAsnSerLys....... 286

    1124 TTTATCATGCTGCAGGTGGGGTCAACAGTTATCGACCCAGACTG...... 1167
                               :::::::|||         ||||||
     287 ....................IleGlySerThrAlaValArgLeuAlaGly 296

    1168 ..............AATAATGGAGAAAATATTTCCTTTATTGACAAAGG 1202
                       |||||||||:::|||::::::|||   |||
     297 AsnGluLysAspAlaAsnAsnGlyGlnAsnValThrPheGluAsp..... 311

    1203 AAAAGGTGAATTGATACTTACCAGCAACATCAACCAAGGCGCGGGCGGTT 1252
          :::|||   |||:::|||   :::||||||||||||||||||||||||
     312 .AsnGlyThrLeuValLeuAspGlnAsnIleAsnGlnGlyAlaGlyGlyL 328

    1253 TGTATTTTGAGGGTAATTTTACGGTCTCGCCTAAAAACAACGAA...ACG 1299
         ||:::|||:::|||:::::::||||||:::      ||||||:::   |||
     328 euPhePheLysGlyAspTyrThrValLysGlyIleAsnAsnAspIleThr 344

    1300 TGGCAAGGCGCGGGCGTTCATATCAGTGATGGCAGTACCGTTACTTGGAA 1349
         |||   ||||||||||:::   ::::::||||||:::   |||:::|||::
     345 TrpLeuGlyAlaGlyIleAspValThrAspGlyLysLysValValTrpGl 361

    1350 AGTAAACGGCGTGGCAAACGACCGCCTGTCCAAAATCGGCAAAGGCACGC 1399
         :|||:::::::      :::||||||||:::||||||||||||||||||
     361 nValLysAsnProAsnGlyAspArgLeuAlaLysIleGlyLysGlyThrL 378

    1400 TGCTGGTTCAAGCCAAAGGGGAAAACCAAGGCTCGGTCAGCGTGGGCGAC 1449
         ||   :::::::::   |||   |||||||||:::::::::|||||||||
     378 euGluIleAsnGlyThrGlyValAsnGlnGlyGlnLeuLysValGlyAsp 394

    1450 GGTAAAGTCATCTTAGATCAGCAGGCGGACGATCAAGGCAAAAAACAAGC 1499
         |||   ||||||||||:::||||||||||||   :::   |||   |||||
     395 GlyThrValIleLeuAsnGlnGlnAlaAspAlaAspLysLysValGlnAl 411

    1500 CTTTAGTGAAATCGGCTTGGTCAGCGGCAGGGGGACGGTGCAACTGAATG 1549
         |||||||:::::|||:::||||||||||||||||||:::   ||||||:
```

```
     411 aPheSerGlnValGlyIleValSerGlyArgGlyThrLeuValLeuAsnS 428

    1550 CCGATAATCAGTTCAACCCCGACAAACTCTATTTCGGCTTTCGCGGCGGA 1599
         :::::||||||:::||||||||:::||||||||||||||||||||||||
     428 erSerAsnGlnIleAsnProAspAsnLeuTyrPheGlyPheArgGlyGly 444

    1600 CGTTTGGATTTGAACGGGCATTCGCTTTCGTTCCACCGCATTCAAAATAC 1649
         |||||||||   ||||||:::::|||:::|||::::::|||:::|||::
     445 ArgLeuAspAlaAsnGlyAsnAspLeuThrPheGluHisIleArgAsnVa 461

    1650 CGATGAAGGGGCGATGATTGTCAACCACAATCAAGACAAAGAATCCACCG 1699
         :|||||||||||||   ||||||||||||||   :::      ||||||:
     461 lAspGluGlyAlaArgIleValAsnHisAsnThrSerHisAlaSerThrI 478

    1700 TTACCATTACAGGCAATAAAGATATTACTACAACCGGCAATAACAACAAC 1749
         ::|||:::||||||::::::   ||||||         |||   |||:::
     478 leThrLeuThrGlyLysSerLeuIleThr.........AsnProAsnSer 491

    1750 TTG 1752
         |||
     492 Leu 492

seq_name: pir2:S61323

seq_documentation_block:
IgA-specific metalloendopeptidase (EC 3.4.24.13) - Neisseria meningitidis (strain HF48 a
N;Alternate names: IgA proteinase; immunoglobulin A1 proteinase
C;Species: Neisseria meningitidis
C;Date: 20-Jul-1996 #sequence_revision 13-Mar-1997 #text_change 08-Dec-2000
C;Accession: S61323; S61324
R;Lomholt, H.; Poulsen, K.; Mogens, K.
Mol. Microbiol. 15, 495-506, 1995
A;Title: Comparative characterization of the iga gene encoding IgA1 protease in Neisseri
A;Reference number: S61314; MUID:95302961
A;Accession: S61323
A;Status: nucleic acid sequence not shown
A;Molecule type: DNA
A;Residues: 1-496 <LOM>
A;Cross-references: EMBL:X82480; NID:g732875; PIDN:CAA57863.1; PID:g732876
A;Experimental source: strain HF48
A;Accession: S61324
A;Status: nucleic acid sequence not shown
A;Molecule type: DNA
A;Residues: 1-496 <LOH>
A;Cross-references: EMBL:X82475; NID:g732879; PIDN:CAA57858.1; PID:g732880
A;Experimental source: strain HF96
C;Genetics:
A;Gene: iga
C;Superfamily: IgA-specific metalloendopeptidase
C;Keywords: hydrolase; metalloproteinase


alignment_scores:
            Quality:  868.00              Length:    544
              Ratio:  2.431                Gaps:     20
 Percent Similarity:  65.625   Percent Identity: 38.603

alignment_block:
 US-09-303-518D-653 x S61323   ..

 Align seg 1/1  to: S61323  from: 1  to: 496

     202 GCGAAAGATATTGAGGTTTACAACAAAAAAGGGGAGTTGGTCGGCAAATC 251
         |||   |||::::::|||   ||||||:::|||:::   :::|||:::::
       1 AlaThrAspLeuSerValLysAsnLysGlnGlyGlnAsnIleGlyAsnAl 17

     252 GATGACGAAAGCCCCGATGATTGATTTTTCTGTGGTATCGCGTAAC...G 298
         ::::::::::::|||||||||||||||||||:::::   |||
      17 aLeuSerAsnValProMetIleAspPheSerValAlaAspValAsnArgA 34

     299 GCGTGGCGGCATTGGCGGGCGATCAATATATTGTGAGCGTGGCACATAAC 348
           :::   ::::::         ||||||   |||||||||   |||
      34 rgThrLeuThrValIleAspProGlnTyrAlaValSerValLysHisVal 50

     349 GGCGGC..............TATAACAATGTTGATTTTGGTGCGGAGGG 383
           |||               |||:::::::   :::         :::
      51 LysGlyAspGluIleSerTyrTyrGlyHisHisAsnGlyHisLeuAspVa 67

     384 AAGCAATCCCGATCAGCACCGCTTTTCTTACCAAATTGTGAAAAGAAATA 433
          ||||||   ::::::         :::|||:::   |||   :::|||:
      67 lSerAsnAspGluAsn.........GluTyrArgSerValAlaGlnAsnA 81

     434 ATTATAAAGCAGGGACTAACGGCCATCCTTATGGCGGC............. 471
         ::|||:::   :::   |||   |||   :::|||:::
      81 spTyrGluProAsnLysAsnTrpHis...HisGlyAsnGlnGlyArgLeu 96

     472 ...GATTATCATATGCCGCGTTTGCACAAATTTGTCACAGATGCAGAACC 518
            ||||||:::|||   ||||||:::||||||||||||||::::::   ||
      97 GluAspTyrAsnMetAlaArgLeuAsnLysPheValThrGluValAlaPr 113

     519 TGTTGAGATGACCAGTTATATGGATGGGTGGAAA...TACGCTGATTTAA 565
         |:::      ||||||         |||   :::   |||   |||   |
     113 oIleAlaProThrSerAlaGlyGlyGlyValGluThrTyrLysAspLysA 130

     566 ATAAATACCCTGATCGTGTTCGAATCGGAGCAGGCAGACAATATTGGCGG 615
         ||::::::   :::   ||||||:::||||||||   |||:::
     130 snArgPheSerGluPheValArgValGlyAlaGlyThrGlnPhe...... 144

     616 TCTGATGAAGACGAACCCAATAACCGCGAAAGTTCATATCATATTGCAAG 665
                 |||   |||:::|||   :::      ::::::::
     145 ...........GluTyrAsnSerArgTyrAsnMetThrGluLeuSerAr 157

     666 CGCATATTCTTGGCTCGTCGGTGGCAATACCTTT.............G 700
          ||||||   :::   ::::::|||:::   :::           :
     157 gAlaTyrArgTyrAlaIleAlaGlyThrProTyrGlnAspValAsnValT 174

     701 CACAAAATGGATCAGGTGGTGGCACAGTCAACTTAGGTAGCGAAAAAATT 750
         :::::|||   :::       |||   :::::::::|||:::::::::
     174 hrSerAsnLeuAsnGlnGluGlyLeuIleGlyPheGlyAspAsnSerLys 190

     751 AAACATAGCCCA.....................TATGGTTTTTTACC 776
           |||||||||                           :::   |||
     191 HisHisSerProGluLysLeuLysGluValLeuSerGlnAsnAlaLeuTh 207

     777 AACAGGAGGCTCATTTGGCGACAGTGGCTCACCAATGTTTATCTATGATG 826
          :::   :::   :::||||||||||||||||||:::|||   ||||||
     207 rAsnTyrAlaValLeuGlyAspSerGlySerProLeuPheAlaTyrAspL 224

     827 CCCAAAAGCAAAAGTGGTTAATTAATGGGGTATTGCAAACAGGCAACCCC 876
           |||:::::::::|||::::::   |||:::   :::
     224 ysGlnGluLysArgTrpValPheLeuGlyAlaTyrAspTyrTrpAlaGly 240

     877 TATATAGGAAAAAGCAATGGCTTCCAGCTAGTTCGTAAAGATTGG...TT 923
         |||       ::::::|||::::::|||             :::|||   ::
     241 Tyr......GlnLysAsnSerTrpGln...........GluTrpAsnIl 251

     924 CTATGATGAAATCTTTGCTGGAGATACCCATTCAGTATTCTACGAACCAC 973
         :|||   :::   ||||||   :::      :::
     251 eTyrLysLysGluPheAlaAspGluIleLysGln.............. 262

     974 ATCAAAATGGGAAATACTTTTTTAACGACAATAATAATGGCGCAGGAAAA 1023
                                 :::||||||         ||||||
     263 ......................ArgAspAsn.........AlaGlyThr 268

    1024 ATCGATGCCAAACATAAACACTATTCTCTACCTTATAGATTAAAAACACG 1073
         |||   :::      :::|||:::                    ||||||
     269 IleLysGlyTyrGlyGluHisHis..............TrpLysThrTh 280

    1074 AACCGTTCAATTGTTTAATGTTTCTTTATCCGAGACAGCAAGAGAACCTG 1123
           :::         |||      :::::::::||||||   :::   :
     280 rGlyThr.........AsnSerHisIleGlySerThrAlaValArgLeuA 294

    1124 TTTATCATGCTGCAGGTGGGGTCAACAGTTATCGACCCAGACTGAATAAT 1173
         ::   :::      |||:::                          ||||||
     294 laGlyAsnGluArgGlyAla.......................AsnAsn 302
```

```
 1174 GGAGAAAATATTTCCTTTATTGACAAAGGAAAAGGTGAATTGATACTTAC 1223
      |||:::|||:::::::|||      ::::::::::|||   |||:::|||
  303 GlyGlnAsnValThrPhe......GluAsnAsnGlyThrLeuValLeuAs 317

 1224 CAGCAACATCAACCAAGGCGCGGGCGGTTTGTATTTTGAGGGTAATTTTA 1273
       :::|||||||||||||||||||||||||||:::|||:::|||:::::::|
  317 pGlnAsnIleAsnGlnGlyAlaGlyGlyLeuPhePheLysGlyAspTyrT 334

 1274 CGGTCTCGCCTAAAAACAACGAA...ACGTGGCAAGGCGCGGGCGTTCAT 1320
      |||||:::      ||||||:::   ||||||   |||||||||:::
  334 hrValLysGlyIleAsnAsnAspIleThrTrpLeuGlyAlaGlyIleAsp 350

 1321 ATCAGTGATGGCAGTACCGTTACTTGGAAAGTAAACGGCGTGGCAAACGA 1370
      ::::::|||||||:::   |||:::|||:::|||::::::      :::||
  351 ValAlaAspGlyLysLysValValTrpGlnValLysAsnProAsnGlyAs 367

 1371 CCGCCTGTCCAAAATCGGCAAAGGCACGCTGCTGGTTCAAGCCAAAGGGG 1420
      |||||||:::||||||||||||||||||||   :::::::::   |||
  367 pArgLeuAlaLysIleGlyLysGlyThrLeuGluIleAsnGlyThrGlyV 384

 1421 AAAACCAAGGCTCGGTCAGCGTGGGCGACGGTAAAGTCATCTTAGATCAG 1470
        |||||||||:::::::::|||||||||||   |||||||||:::|||
  384 alAsnGlnGlyGlnLeuLysValGlyAspGlyThrValIleLeuAsnGln 400

 1471 CAGGCGGACGATCAAGGCAAAAAACAAGCCTTTAGTGAAATCGGCTTGGT 1520
      :::|||||||::::::   |||   |||||||||||||:::::::|||:::||
  401 LysAlaAspSerAsnGlnLysValGlnAlaPheSerGlnValGlyIleVa 417

 1521 CAGCGGCAGGGGGACGGTGCAACTGAATGCCGATAATCAGTTCAACCCCG 1570
      |||||||||||||||:::   ||||||:::::|||||||:::|||||||
  417 lSerGlyArgGlyThrLeuValLeuAsnSerSerAsnGlnIleAsnProA 434

 1571 ACAAACTCTATTTCGGCTTTCGCGGCGGACGTTTGGATTTGAACGGGCAT 1620
      ||:::|||||||||||||||||||||||||||||||   ||||||:::
  434 spAsnLeuTyrPheGlyPheArgGlyGlyArgLeuAspAlaAsnGlyAsn 450

 1621 TCGCTTTCGTTCCACCGCATTCAAAATACCGATGAAGGGGCGATGATTGT 1670
      :::|||:::|||:::::::|||:::|||:::|||||||||||||   |||||
  451 AspLeuThrPheGluHisIleArgAsnValAspGluGlyAlaArgIleVa 467

 1671 CAACCACAATCAAGACAAAGAATCCACCGTTACCATTACAGGCAATAAAG 1720
      ||||||||||            ||||||:::|||:::||||||:::::
  467 lAsnHisAsnThrGlyHisThrSerThrIleThrLeuThrGlyLysSerL 484

 1721 ATATTACTACAACCGGCAATAACAACAACTTG 1752
        ||||||         |||   |||:::|||
  484 euIleThr.........AsnProAsnSerLeu 491
```

seq_name: pir2:S61325

seq_documentation_block:
IgA-specific metalloendopeptidase (EC 3.4.24.13) homolog - Neisseria meningitidis (fragm
C;Species: Neisseria meningitidis
A;Variety: SM1894
C;Date: 23-Jul-1996 #sequence_revision 13-Mar-1997 #text_change 08-Dec-2000
C;Accession: S61325
R;Lomholt, H.; Poulsen, K.; Mogens, K.
Mol. Microbiol. 15, 495-506, 1995
A;Title: Comparative characterization of the iga gene encoding IgA1 protease in Neisseri
A;Reference number: S61314; MUID:95302961
A;Accession: S61325
A;Status: preliminary; nucleic acid sequence not shown
A;Molecule type: DNA
A;Residues: 1-496 <LOM>
A;Cross-references: EMBL:X82476; NID:g732867; PIDN:CAA57859.1; PID:g732868
C;Superfamily: IgA-specific metalloendopeptidase
C;Keywords: hydrolase; metalloproteinase

alignment_scores:
Quality: 860.00 Length: 544
Ratio: 2.416 Gaps: 20
Percent Similarity: 65.441 Percent Identity: 38.419

alignment_block:
US-09-303-518D-653 x S61325 ..

Align seg 1/1 to: S61325 from: 1 to: 496

```
  202 GCGAAAGATATTGAGGTTTACAACAAAAAAGGGGAGTTGGTCGGCAAATC 251
      |||   |||::::::|||   ||||||:::|||:::   :::|||:::::
    1 AlaThrAspLeuSerValLysAsnLysGlnGlyGlnAsnIleGlyAsnAl 17

  252 GATGACGAAAGCCCCGATGATTGATTTTTCTGTGGTATCGCGTAAC...G 298
      :::::::::::::|||||||||||||||||||||::::::   |||
   17 aLeuSerAsnValProMetIleAspPheSerValAlaAspValAsnArgA 34

  299 GCGTGGCGGCATTGGCGGGCGATCAATATATTGTGAGCGTGGCACATAAC 348
        :::   :::::          ||||||   |||||||||   |||
   34 rgThrLeuThrValIleAspProGlnTyrAlaValSerValLysHisVal 50

  349 GGCGGC..............TATAACAATGTTGATTTTGGTGCGGAGGG 383
        |||               |||::::::   :::       :::
   51 LysGlyAspGluIleSerTyrTyrGlyHisHisAsnGlyHisLeuAspVa 67

  384 AAGCAATCCCGATCAGCACCGCTTTTCTTACCAAATTGTGAAAAGAAATA 433
       ||||||   ::::::          :::|||:::   |||   :::|||:
   67 lSerAsnAspGluAsn.........GluTyrArgSerValAlaGlnAsnA 81

  434 ATTATAAAGCAGGGACTAACGGCCATCCTTATGGCGGC............. 471
      ::||||:::   :::   |||   |||   :::|||:::
   81 spTyrGluProAsnLysAsnTrpHis...HisGlyAsnGlnGlyArgLeu 96

  472 ...GATTATCATATGCCGCGTTTGCACAAATTTGTCACAGATGCAGAACC 518
         ||||||:::|||   ||||||:::||||||||||||:::::::   ||
   97 GluAspTyrAsnMetAlaArgLeuAsnLysPheValThrGluValAlaPr 113

  519 TGTTGAGATGACCAGTTATATGGATGGGTGGAAA...TACGCTGATTTAA 565
      |:::      ||||||         |||   :::   |||   |||   |
  113 oIleAlaProThrSerAlaGlyGlyGlyValGluThrTyrLysAspLysA 130

  566 ATAAATACCCTGATCGTGTTCGAATCGGAGCAGGCAGACAATATTGGCGG 615
      ||::::::   :::   ||||||:::||||||||   |||:::
  130 snArgPheSerGluPheValArgValGlyAlaGlyThrGlnPhe...... 144

  616 TCTGATGAAGACGAACCCAATAACCGCGAAAGTTCATATCATATTGCAAG 665
                |||   |||:::|||   :::       :::::::::
  145 ............GluTyrAsnSerArgTyrAsnMetThrGluLeuSerAr 157

  666 CGCATATTCTTGGCTCGTCGGTGGCAATACCTTT..............G 700
       ||||||   :::   ::::::|||:::   :::             :
  157 gAlaTyrArgTyrAlaIleAlaGlyThrProTyrGlnAspValAsnValT 174

  701 CACAAAATGGATCAGGTGGTGGCACAGTCAACTTAGGTAGCGAAAAAATT 750
      ::::::|||   :::       |||   ::::::::::|||:::::::::
  174 hrSerAsnLeuAsnGlnGluGlyLeuIleGlyPheGlyAspAsnSerLys 190

  751 AAACATAGCCCA.....................TATGGTTTTTTACC 776
         |||||||||                     :::   |||
  191 HisHisSerProGluLysLeuLysGluValLeuSerGlnAsnAlaLeuTh 207

  777 AACAGGAGGCTCATTTGGCGACAGTGGCTCACCAATGTTTATCTATGATG 826
       :::   :::   :::|||||||||||||||||||:::|||   ||||||
  207 rAsnTyrAlaValLeuGlyAspSerGlySerProLeuPheAlaTyrAspL 224

  827 CCCAAAAGCAAAAGTGGTTAATTAATGGGGTATTGCAAACAGGCAACCCC 876
        |||:::::::::|||::::::   |||:::   :::
  224 ysGlnGluLysArgTrpValPheLeuGlyAlaTyrAspTyrTrpAlaGly 240

  877 TATATAGGAAAAAGCAATGGCTTCCAGCTAGTTCGTAAAGATTGG...TT 923
      |||       ::::::|||:::::::|||            :::|||   ::
  241 Tyr......GlnLysAsnSerTrpGln............GluTrpAsnIl 251

  924 CTATGATGAAATCTTTGCTGGAGATACCCATTCAGTATTCTACGAACCAC 973
      :|||   :::   ||||||   :::      :::
  251 eTyrLysLysGluPheAlaAspGluIleLysGln..............  262
```

```
 974 ATCAAAATGGGAAATACTTTTTTAACGACAATAATAATGGCGCAGGAAAA 1023
                          :::||||||         ||||||
 263 .....................ArgAspAsn.........AlaGlyThr 268

1024 ATCGATGCCAAACATAAACACTATTCTCTACCTTATAGATTAAAAACACG 1073
     |||   :::      :::|||:::                 ||||||
 269 IleLysGlyTyrGlyGluHisHis..............TrpLysThrTh 280

1074 AACCGTTCAATTGTTTAATGTTTCTTTATCCGAGACAGCAAGAGAACCTG 1123
       :::         |||      :::::::::||||||   :::   :
 280 rGlyThr.........AsnSerHisIleGlySerThrAlaValArgLeuA 294

1124 TTTATCATGCTGCAGGTGGGGTCAACAGTTATCGACCCAGACTGAATAAT 1173
     ::   :::      |||:::                   ||||||
 294 laGlyAsnGluArgGlyAla.....................AsnAsn 302

1174 GGAGAAAATATTTCCTTTATTGACAAAGGAAAAGGTGAATTGATACTTAC 1223
     |||:::|||::::::|||      ::::::::::|||   |||:::|||
 303 GlyGlnAsnValThrPhe......GluAsnAsnGlyThrLeuValLeuAs 317

1224 CAGCAACATCAACCAAGGCGCGGGCGGTTTGTATTTTGAGGGTAATTTTA 1273
      :::||||||||||||||||||||||||||:::|||:::|||:::::::|
 317 pGlnAsnIleAsnGlnGlyAlaGlyGlyLeuPhePheLysGlyAspTyrT 334

1274 CGGTCTCGCCTAAAAACAACGAA...ACGTGGCAAGGCGCGGGCGTTCAT 1320
     |||||:::      ||||||:::   ||||||   |||||||||:::
 334 hrValLysGlyIleAsnAsnAspIleThrTrpLeuGlyAlaGlyIleAsp 350

1321 ATCAGTGATGGCAGTACCGTTACTTGGAAAGTAAACGGCGTGGCAAACGA 1370
     ::::::|||||:::   |||:::|||:::|||::::::      :::||
 351 ValAlaAspGlyLysLysValValTrpGlnValLysAsnProAsnGlyAs 367

1371 CCGCCTGTCCAAAATCGGCAAAGGCACGCTGCTGGTTCAAGCCAAAGGGG 1420
     |||||||:::|||||||||||||||||||||   :::::::::   |||
 367 pArgLeuAlaLysIleGlyLysGlyThrLeuGluIleAsnGlyThrGlyV 384

1421 AAAACCAAGGCTCGGTCAGCGTGGGCGACGGTAAAGTCATCTTAGATCAG 1470
       |||||||||:::::::::|||||||||||   |||||||||:::|||
 384 alAsnGlnGlyGlnLeuLysValGlyAspGlyThrValIleLeuAsnGln 400

1471 CAGGCGGACGATCAAGGCAAAAAACAAGCCTTTAGTGAAATCGGCTTGGT 1520
     :::||||||::::::   |||   ||||||||||||:::::|||:::||
 401 LysAlaAspSerAsnGlnLysValGlnAlaPheSerGlnValGlyIleVa 417

1521 CAGCGGCAGGGGGACGGTGCAACTGAATGCCGATAATCAGTTCAACCCCG 1570
     ||||||||||   |||:::   ||||||::::::||||||:::|||||||
 417 lSerGlyArgArgThrLeuValLeuAsnSerSerAsnGlnIleAsnProA 434

1571 ACAAACTCTATTTCGGCTTTCGCGGCGGACGTTTGGATTTGAACGGGCAT 1620
     ||:::|||||||||||||||||||||||||||||||||-  ||||||:::
 434 spAsnLeuTyrPheGlyPheArgGlyGlyArgLeuAspAlaAsnGlyAsn 450

1621 TCGCTTTCGTTCCACCGCATTCAAAATACCGATGAAGGGGCGATGATTGT 1670
     :::|||:::|||:::::::|||:::|||:::||||||||||||   |||||
 451 AspLeuThrPheGluHisIleArgAsnValAspGluGlyAlaArgIleVa 467

1671 CAACCACAATCAAGACAAAGAATCCACCGTTACCATTACAGGCAATAAAG 1720
     ||||||||||            ||||||:::|||:::||||||::::::
 467 lAsnHisAsnThrGlyHisThrSerThrIleThrLeuThrGlyLysSerL 484

1721 ATATTACTACAACCGGCAATAACAACAACTTG 1752
       ||||||         |||   |||:::|||
 484 euIleThr.........AsnProAsnSerLeu 491
```

seq_name: pir2:S61329

seq_documentation_block:
IgA-specific metalloendopeptidase (EC 3.4.24.13) homolog - Neisseria meningitidis (fragm
C;Species: Neisseria meningitidis
A;Variety: NGC16
C;Date: 23-Jul-1996 #sequence_revision 13-Mar-1997 #text_change 08-Dec-2000
C;Accession: S61329

R;Lomholt, H.; Poulsen, K.; Mogens, K.
Mol. Microbiol. 15, 495-506, 1995
A;Title: Comparative characterization of the iga gene encoding IgA1 protease in Neiss
A;Reference number: S61314; MUID:95302961
A;Accession: S61329
A;Status: preliminary; nucleic acid sequence not shown
A;Molecule type: DNA
A;Residues: 1-508 <LOM>
A;Cross-references: EMBL:X82485; NID:g732869; PIDN:CAA57868.1; PID:g732870
C;Superfamily: IgA-specific metalloendopeptidase
C;Keywords: hydrolase; metalloproteinase

alignment_scores:
Quality: 859.50 Length: 547
Ratio: 2.463 Gaps: 18
Percent Similarity: 63.803 Percent Identity: 38.391

alignment_block:
US-09-303-518D-653 x S61329 ..

Align seg 1/1 to: S61329 from: 1 to: 508

```
 202 GCGAAAGATATTGAGGTTTACAACAAAAAAGGGGAGTTGGTCGGCAAATC 251
     |||   |||::::::|||   ||||||:::|||:::   :::|||:::::
   1 AlaThrAspLeuSerValLysAsnLysGlnGlyGlnAsnIleGlyAsnAl 17

 252 GATGACGAAAGCCCCGATGATTGATTTTTCTGTGGTATCGCGTAAC...G 298
     ::::::::::::|||||||||||||||||||||||:::::   |||
  17 aLeuSerAsnValProMetIleAspPheSerValAlaAspValAsnLysA 34

 299 GCGTGGCGGCATTGGCGGGCGATCAATATATTGTGAGCGTGGCACAT... 345
       :::|||:::::::::      ||||||   |||||||||   |||
  34 rgIleAlaThrValValAspProGlnTyrAlaValSerValLysHisAla 50

 346 .......................AACGGCGGCTATAACAATGTTGATTT 371
                                 |||   ||||||:::   :::
  51 LysAlaGluValHisThrPheTyrTyrGlyGlnTyrAsnGlyHisAsnAs 67

 372 TGGTGCGGAGGGAAGCAATCCCGATCAGCACCGCTTTTCTTACCAAATTG 421
         |||:::   :::|||                  :::|||::::::|
  67 pValAlaAspLysGluAsn.................GluTyrArgValV 78

 422 TGAAAAGAAATAATTATAAAGCAGGGACTAACGGCCATCCTTATGGCGGC 471
     ||::::::|||||||||:::                |||      |||
  78 alGluGlnAsnAsnTyrGluPro............HisLysAlaTrpGly 90

 472 .......................GATTATCATATGCCGCGTTTGCACAA 497
                              |||||||:::|||   |||::::::||
  91 AlaSerAsnLeuGlyArgLeuGluAspTyrAsnMetAlaArgPheAsnLy 107

 498 ATTTGTCACAGATGCAGAACCTGTTGAGATGACCAGTTATATGGATGGGT 547
     ||||||||||::::::   |||:::      |||:::          |||
 107 sPheValThrGluValAlaProIleAlaProThrAspAlaGlyGlyGlyL 124

 548 GG...AAATACGCTGATTTAAATAAATACCCTGATCGTGTTCGAATCGGA 594
            |||   |||   |||::::::   :::   ||||||:::|||
 124 euAspThrTyrLysAspLysAsnArgPheSerSerPheValArgValGly 140

 595 GCAGGCAGACAA......TATTGGCGGTCTGATGAAGACGAACCCAATAA 638
     ||||||||||||      |||  ::::::        :::   :::||
 141 AlaGlyArgGlnLeuValTyrGluLysGlyAlaTyrHisGlnGluGlyAs 157

 639 CCGCGAAAGTTCATAT.........CATATTGCAAGCGCATATTCTTGGC 679
     |   |||:::::::|||           :::::::::||||||   :::
 157 n...GluLysGlyTyrAspLeuArgAspLeuSerGlnAlaTyrArgTyrA 173

 680 TCGTCGGTGGC.................................AAT 693
       :::::::|||                                 |||
 173 laIleAlaGlyThrProTyrLysAspIleAsnIleAspGlnThrMetAsn 189

 694 ACCTTTGCACAAAATGGATCAGGTGGTGGCACAGTCAACTTAGGTAGCGA 743
     |||   :::      |||   |||:::   :::::::::   :::::::||
```

```
 190 ThrGluGlyLeuIleGlyPheGlyAsnHisAsnThrHisTyrSerAlaGl 206

 744 AAAAATTAAACATAGC...CCATATGGTTTTTTACCAACAGGAGGCTCAT 790
     |:::::::|||::::::               |||   :::   |||   :
 206 uGluLeuLysGlnAlaLeuSerGlnAspAlaLeuThrAsnTyrGlyValL 223

 791 TTGGCGACAGTGGCTCACCAATGTTTATCTATGATGCCCAAAAGCAAAAG 840
     ::|||||||||||||||||:::|||   :::|||   ||||||::::::
 223 euGlyAspSerGlySerProLeuPheAlaPheAspLysGlnLysAsnGln 239

 841 TGGTTAATTAATGGGGTATTGCAAACAGGCAACCCCTATATAGGAAAAAG 890
     |||::::::   |||:::   :::             |||      |||||
 240 TrpValPheLeuGlyThrTyrAspTyrTrpAlaGlyTyrGlyLysLysSe 256

 891 CAATGGCTTCCAGCTAGTTCGTAAAGATTGGTTCTATGATGAAATCTTTG 940
     |                ::::   |||      |||   |||:::|||
 256 rTrpGlnGluTrpAsnIleTyrLysLysGluPheAlaAspLysIleLysG 273

 941 CTGGAGATACCCATTCAGTATTCTACGAACCACATCAAAATGGGAAATAC 990
         |||
 273 lnArgAsp..................................... 275

 991 TTTTTTAACGACAATAATAATGGCGCAGGAAAAATCGATGCCAAACATAA 1040
                  |||   |||:::   |||   ::::::::::|||
 276 ..............AsnAlaGlyThrIleLysGlyAsnGlyGluHis.. 286

1041 ACACTATTCTCTACCTTATAGATTAAAAACACGAACCGTTCAATTGTTTA 1090
      |||::::::::::   :::                          |
 287 .HisTrpAsnIleThrPheGlyThr.....................A 295

1091 ATGTTTCTTTATCCGAGACAGCAAGAGAACCTGTTTATCATGCTGCAGGT 1140
     ||      ::::::::::||||||               :::   ||||||
 295 snSerHisIleGlySerThrAla...........ValArgLeuAlaGly 307

1141 GGGGTCAACAGTTATCGACCCAGACTGAATAATGGAGAAAATATTTCCTT 1190
     :::   ::::::              |||||||||:::|||:::::::||
 308 AsnGluArgAsp...........AlaAsnAsnGlyGlnAsnValThrPh 320

1191 TATTGACAAAGGAAAAGGTGAATTGATACTTACCAGCAACATCAACCAAG 1240
     |   |||      :::|||   |||:::|||   :::|||||||||||||
 320 eGluAsp......AsnGlyThrLeuValLeuAspGlnAsnIleAsnGlnG 335

1241 GCGCGGGCGGTTTGTATTTTGAGGGTAATTTTACGGTCTCGCCTAAAAAC 1290
     |||||||||||||:::|||:::|||:::::|||||||:::      |||
 335 lyAlaGlyGlyLeuPhePheLysGlyAspTyrThrValLysGlyAlaAsn 351

1291 AACGAA...ACGTGGCAAGGCGCGGGCGTTCATATCAGTGATGGCAGTAC 1337
     |||:::   ||||||   |||||||||:::   :::::|||||:::
 352 AsnAspIleThrTrpLeuGlyAlaGlyIleAspValAlaAspGlyLysLy 368

1338 CGTTACTTGGAAAGTAAACGGCGTGGCAAACGACCGCCTGTCCAAAATCG 1387
      |||:::|||:::|||:::::      :::|||||||||:::|||||||
 368 sValValTrpGlnValLysAsnProAsnGlyAspArgLeuAlaLysIleG 385

1388 GCAAAGGCACGCTGCTGGTTCAAGCCAAAGGGGAAAACCAAGGCTCGGTC 1437
     |||||||||||||   :::::::::   |||   |||||||||:::::
 385 lyLysGlyThrLeuGluIleAsnGlyThrGlyValAsnGlnGlyGlnLeu 401

1438 AGCGTGGGCGACGGTAAAGTCATCTTAGATCAGCAGGCGGACGATCAAGG 1487
     :::|||||||||||||   |||||||||:::|||:::|||||:::::::
 402 LysValGlyAspGlyThrValIleLeuAsnGlnLysAlaAspSerAsnGl 418

1488 CAAAAAACAAGCCTTTAGTGAAATCGGCTTGGTCAGCGGCAGGGGGACGG 1537
      |||   ||||||||||||:::::|||:::|||||||||||||||||:
 418 nLysValGlnAlaPheSerGlnValGlyIleValSerGlyArgGlyThrL 435

1538 TGCAACTGAATGCCGATAATCAGTTCAACCCCGACAAACTCTATTTCGGC 1587
     ::   ||||||:::::|||||||:::||||||||||:::||||||||||
 435 euValLeuAsnSerSerAsnGlnIleAsnProAspAsnLeuTyrPheGly 451

1588 TTTCGCGGCGGACGTTTGGATTTGAACGGGCATTCGCTTTCGTTCCACCG 1637
     ||||||||||||||||||||   ||||||:::::|||:::|||:::::
 452 PheArgGlyGlyArgLeuAspAlaAsnGlyAsnAspLeuThrPheGluHi 468

1638 CATTCAAAATACCGATGAAGGGGCGATGATTGTCAACCACAATCAAGACA 1687
     :|||:::|||:::||||||||||||   ||||||||||||||||
 468 sIleArgAsnValAspGluGlyAlaArgIleValAsnHisAsnThrGlyH 485

1688 AAGAATCCACCGTTACCATTACAGGCAATAAAGATATTACT 1728
          ||||||:::|||:::||||||::::::   ||||||
 485 isAlaSerThrIleThrLeuThrGlyLysSerLeuIleThr 498
```

seq_name: pir2:S61322

seq_documentation_block:
IgA-specific metalloendopeptidase (EC 3.4.24.13) homolog - Neisseria meningitidis (fr
C;Species: Neisseria meningitidis
A;Variety: HF54
C;Date: 20-Jul-1996 #sequence_revision 13-Mar-1997 #text_change 08-Dec-2000
C;Accession: S61322
R;Lomholt, H.; Poulsen, K.; Mogens, K.
Mol. Microbiol. 15, 495-506, 1995
A;Title: Comparative characterization of the iga gene encoding IgA1 protease in Neiss
A;Reference number: S61314; MUID:95302961
A;Accession: S61322
A;Status: preliminary; nucleic acid sequence not shown
A;Molecule type: DNA
A;Residues: 1-496 <LOM>
A;Cross-references: EMBL:X82473; NID:g732877; PIDN:CAA57856.1; PID:g732878
C;Superfamily: IgA-specific metalloendopeptidase
C;Keywords: hydrolase; metalloproteinase

alignment_scores:
Quality: 858.50 Length: 540
Ratio: 2.432 Gaps: 19
Percent Similarity: 65.370 Percent Identity: 38.333

alignment_block:
US-09-303-518D-653 x S61322 ..

Align seg 1/1 to: S61322 from: 1 to: 496

```
 202 GCGAAAGATATTGAGGTTTACAACAAAAAAGGGGAGTTGGTCGGCAAATC 251
     |||   |||::::::|||   ||||||:::|||:::   :::|||:::::
   1 AlaThrAspLeuSerValLysAsnLysGlnGlyGlnAsnIleGlyAsnAl 17

 252 GATGACGAAAGCCCCGATGATTGATTTTTCTGTGGTATCGCGTAAC...G 298
     ::::::::::::|||||||||||||||||||||::::::   |||
  17 aLeuSerAsnValProMetIleAspPheSerValAlaAspValAsnArgA 34

 299 GCGTGGCGGCATTGGCGGGCGATCAATATATTGTGAGCGTGGCACATAAC 348
       :::   ::::::          ||||||   |||||||||   |||
  34 rgThrLeuThrValIleAspProGlnTyrAlaValSerValLysHisVal 50

 349 GGCGGC..............TATAACAATGTTGATTTTGGTGCGGAGGG 383
       |||                |||::::::   :::          :::
  51 LysGlyAspGluIleSerTyrTyrGlyHisHisAsnGlyHisLeuAspVa 67

 384 AAGCAATCCCGATCAGCACCGCTTTTCTTACCAAATTGTGAAAAGAAATA 433
      ||||||   ::::::          :::|||:::   |||   :::|||:
  67 lSerAsnAspGluAsn.........GluTyrArgSerValAlaGlnAsnA 81

 434 ATTATAAAGCAGGGACTAACGGCCATCCTTATGGCGGC............ 471
     ::|||:::   :::   |||   |||   :::|||:::
  81 spTyrGluProAsnLysAsnTrpHis...HisGlyAsnGlnGlyArgLeu 96

 472 ...GATTATCATATGCCGCGTTTGCACAAATTTGTCACAGATGCAGAACC 518
       |||||||:::|||   ||||||:::||||||||||||::::::   ||
  97 GluAspTyrAsnMetAlaArgLeuAsnLysPheValThrGluValAlaPr 113

 519 TGTTGAGATGACCAGTTATATGGATGGGTGGAAA...TACGCTGATTTAA 565
     |:::      ||||||         |||   :::   |||   |||   |
 113 oIleAlaProThrSerAlaGlyGlyGlyValGluThrTyrLysAspLysA 130

 566 ATAAATACCCTGATCGTGTTCGAATCGGAGCAGGCAGACAATATTGGCGG 615
```

```
     ||:::::::   :::   |||||||:::|||||||||   |||:::
 130 snArgPheSerGluPheValArgValGlyAlaGlyThrGlnPhe..... 144

 616 TCTGATGAAGACGAACCCAATAACCGCGAAAGTTCATATCATATTGCAAG 665
                 |||   |||:::|||   :::      :::::::::
 145 ............GluTyrAsnSerArgTyrAsnMetThrGluLeuSerAr 157

 666 CGCATATTCTTGGCTCGTCGGTGGCAATACCTTT..............G 700
      ||||||   :::   ::::::|||:::   :::             :
 157 gAlaTyrArgTyrAlaIleAlaGlyThrProTyrGlnAspValAsnValT 174

 701 CACAAAATGGATCAGGTGGTGGCACAGTCAACTTAGGTAGCGAAAAAATT 750
     :::::|||   :::      |||   :::::::::|||:::::::::
 174 hrSerAsnLeuAsnGlnGluGlyLeuIleGlyPheGlyAspAsnSerLys 190

 751 AAACATAGCCCA.......................TATGGTTTTTTACC 776
        |||||||||                       :::   |||
 191 HisHisSerProGluLysLeuLysGluValLeuSerGlnAsnAlaLeuTh 207

 777 AACAGGAGGCTCATTTGGCGACAGTGGCTCACCAATGTTTATCTATGATG 826
      :::   :::   :::||||||||||||||||||||:::|||   ||||||
 207 rAsnTyrAlaValLeuGlyAspSerGlySerProLeuPheAlaTyrAspL 224

 827 CCCAAAAGCAAAAGTGGTTAATTAATGGGGTATTGCAAACAGGCAACCCC 876
       |||:::::::::|||::::::   |||:::   :::
 224 ysGlnGluLysArgTrpValPheLeuGlyAlaTyrAspTyrTrpAlaGly 240

 877 TATATAGGAAAAAGCAATGGCTTCCAGCTAGTTCGTAAAGATTGG...TT 923
     |||      ::::::|||::::::|||            :::|||   ::
 241 Tyr......GlnLysAsnSerTrpGln............GluTrpAsnIl 251

 924 CTATGATGAAATCTTTGCTGGAGATACCCATTCAGTATTCTACGAACCAC 973
     :|||   :::   ||||||   :::     :::
 251 eTyrLysLysGluPheAlaAspGluIleLysGln..............  262

 974 ATCAAAATGGGAAATACTTTTTTAACGACAATAATAATGGCGCAGGAAAA 1023
                           :::|||||||         ||||||
 263 ......................ArgAspAsn.........AlaGlyThr 268

1024 ATCGATGCCAAACATAAACACTATTCTCTACCTTATAGATTAAAAACACG 1073
     |||   :::      :::|||:::                   ||||||
 269 IleLysGlyTyrGlyGluHisHis..............TrpLysThrTh 280

1074 AACCGTTCAATTGTTTAATGTTTCTTTATCCGAGACAGCAAGAGAACCTG 1123
         :::          |||      :::::::::||||||   :::   :
 280 rGlyThr.........AsnSerHisIleGlySerThrAlaValArgLeuA 294

1124 TTTATCATGCTGCAGGTGGGGTCAACAGTTATCGACCCAGACTGAATAAT 1173
     ::   :::      |||:::                        ||||||
 294 laGlyAsnGluArgGlyAla.......................AsnAsn 302

1174 GGAGAAAATATTTCCTTTATTGACAAAGGAAAAGGTGAATTGATACTTAC 1223
     |||::::|||:::::|||      :::::::::|||   |||:::|||
 303 GlyGlnAsnValThrPhe......GluAsnAsnGlyThrLeuValLeuAs 317

1224 CAGCAACATCAACCAAGGCGCGGGCGGTTTGTATTTTGAGGGTAATTTTA 1273
      :::||||||||||||||||||||||||||:::|||:::|||:::::::|
 317 pGlnAsnIleAsnGlnGlyAlaGlyGlyLeuPhePheLysGlyAspTyrT 334

1274 CGGTCTCGCCTAAAAACAACGAA...ACGTGGCAAGGCGCGGGCGTTCAT 1320
     |||||:::      ||||||      ||||||   |||||||||:::
 334 hrValLysGlyAlaAsnAsnGlyIleThrTrpLeuGlyAlaGlyIleAsp 350

1321 ATCAGTGATGGCAGTACCGTTACTTGGAAAGTAAACGGCGTGGCAAACGA 1370
     :::::|||||:::   |||:::|||:::|||:::::      :::||
 351 ValAlaAspGlyLysLysValValTrpGlnValLysAsnProAsnGlyAs 367

1371 CCGCCTGTCCAAAATCGGCAAAGGCACGCTGCTGGTTCAAGCCAAAGGGG 1420
     |||||||:::|||||||||||||||||||||   :::::::::   |||
 367 pArgLeuAlaLysIleGlyLysGlyThrLeuGluIleAsnGlyThrGlyV 384

1421 AAAACCAAGGCTCGGTCAGCGTGGGCGACGGTAAAGTCATCTTAGATCAG 1470
       |||||||||:::::::::|||||||||||||   |||||||||:::|||

 384 alAsnGlnGlyGlnLeuLysValGlyAspGlyThrValIleLeuAsnGln 400

1471 CAGGCGGACGATCAAGGCAAAAAACAAGCCTTTAGTGAAATCGGCTTGGT 1520
     |||||||||   :::   |||   |||||||||||||:::::|||:::||
 401 GlnAlaAspAlaAspLysLysValGlnAlaPheSerGlnValGlyIleVa 417

1521 CAGCGGCAGGGGGACGGTGCAACTGAATGCCGATAATCAGTTCAACCCCG 1570
     ||||||||||   |||:::   ||||||:::   :::|||:::||||||:
 417 lSerGlyArgArgThrLeuValLeuAsnSerProAspGlnIleAsnProA 434

1571 ACAAACTCTATTTCGGCTTTCGCGGCGGACGTTTGGATTTGAACGGGCAT 1620
     :::::||||||||||||||||||||||||||||||||   ||||||:::
 434 snAsnLeuTyrPheGlyPheArgGlyGlyArgLeuAspAlaAsnGlyAsn 450

1621 TCGCTTTCGTTCCACCGCATTCAAAATACCGATGAAGGGGCGATGATTGT 1670
     :::|||:::|||:::::|||:::|||:::||||||||||||   |||||
 451 AspLeuThrPheGluHisIleArgAsnValAspGluGlyAlaArgIleVa 467

1671 CAACCACAATCAAGACAAAGAATCCACCGTTACCATTACAGGCAATAAAG 1720
     ||||||||||   |||:::   ||||||:::|||:::||||||:::::
 467 lAsnHisAsnThrAspArgAlaSerThrIleThrLeuThrGlyLysSerL 484

1721 ATATTACTACAACCGGCAAT 1740
       ||||||:::      |||
 484 euIleThrAlaProGlnAsn 490
```

seq_name: pir2:S61327

seq_documentation_block:
IgA-specific metalloendopeptidase (EC 3.4.24.13) homolog - Neisseria meningitidis (fr
C;Species: Neisseria meningitidis
A;Variety: NCG80
C;Date: 23-Jul-1996 #sequence_revision 13-Mar-1997 #text_change 08-Dec-2000
C;Accession: S61327
R;Lomholt, H.; Poulsen, K.; Mogens, K.
Mol. Microbiol. 15, 495-506, 1995
A;Title: Comparative characterization of the iga gene encoding IgA1 protease in Neiss
A;Reference number: S61314; MUID:95302961
A;Accession: S61327
A;Status: preliminary; nucleic acid sequence not shown
A;Molecule type: DNA
A;Residues: 1-496 <LOM>
A;Cross-references: EMBL:X82479; NID:g732883; PIDN:CAA57862.1; PID:g732884
C;Superfamily: IgA-specific metalloendopeptidase
C;Keywords: hydrolase; metalloproteinase

alignment_scores:
Quality: 856.50 Length: 536
Ratio: 2.433 Gaps: 20
Percent Similarity: 65.672 Percent Identity: 38.806

alignment_block:
US-09-303-518D-653 x S61327 ..

Align seg 1/1 to: S61327 from: 1 to: 496

```
  202 GCGAAAGATATTGAGGTTTACAACAAAAAAGGGGAGTTGGTCGGCAAATC 251
      |||   |||:::::|||   ||||||:::|||:::   :::|||:::::
    1 AlaThrAspLeuSerValLysAsnLysGlnGlyGlnAsnIleGlyAsnAl 17

  252 GATGACGAAAGCCCCGATGATTGATTTTTCTGTGGTATCGCGTAAC...G 298
      :::::::::::::|||||||||||||||||||||||:::::   |||
   17 aLeuSerAsnValProMetIleAspPheSerValAlaAspValAsnArgA 34

  299 GCGTGGCGGCATTGGCGGGCGATCAATATATTGTGAGCGTGGCACATAAC 348
        :::   ::::::         ||||||   |||||||||   |||
   34 rgThrLeuThrValIleAspProGlnTyrAlaValSerValLysHisVal 50

  349 GGCGGC..:.............TATAACAATGTTGATTTTGGTGCGGAGGG 383
         |||               |||::::::   :::          :::
   51 LysGlyAspGluIleSerTyrTyrGlyHisHisAsnGlyHisLeuAspVa 67
```

```
 384 AAGCAATCCCGATCAGCACCGCTTTTCTTACCAAATTGTGAAAAGAAATA 433
     ||||||  ::::::         :::|||:::  |||  :::|||:
  67 lSerAsnAspGluAsn.........GluTyrArgSerValAlaGlnAsnA 81

 434 ATTATAAAGCAGGGACTAACGGCCATCCTTATGGCGGC............ 471
     ::|||:::  :::  |||  |||  :::|||:::
  81 spTyrGluProAsnLysAsnTrpHis...HisGlyAsnGlnGlyArgLeu 96

 472 ...GATTATCATATGCCGCGTTTGCACAAATTTGTCACAGATGCAGAACC 518
        ||||||:::|||  ||||||::||||||||||||:::::::  ||
  97 GluAspTyrAsnMetAlaArgLeuAsnLysPheValThrGluValAlaPr 113

 519 TGTTGAGATGACCAGTTATATGGATGGGTGGAAA...TACGCTGATTTAA 565
     |:::      ||||||         |||  :::  |||  |||  |
 113 oIleAlaProThrSerAlaGlyGlyGlyValGluThrTyrLysAspLysA 130

 566 ATAAATACCCTGATCGTGTTCGAATCGGAGCAGGCAGACAATATTGGCGG 615
     ||::::::  :::  ||||||:::||||||||  |||:::
 130 snArgPheSerGluPheValArgValGlyAlaGlyThrGlnPhe..... 144

 616 TCTGATGAAGACGAACCCAATAACCGCGAAAGTTCATATCATATTGCAAG 665
               |||  |||:::|||  :::     :::::::::
 145 ...........GluTyrAsnSerArgTyrAsnMetThrGluLeuSerAr 157

 666 CGCATATTCTTGGCTCGTCGGTGGCAATACCTTT.............G 700
     ||||||  :::  ::::::|||:::  :::             :
 157 gAlaTyrArgTyrAlaIleAlaGlyThrProTyrGlnAspValAsnValT 174

 701 CACAAAATGGATCAGGTGGTGGCACAGTCAACTTAGGTAGCGAAAAAATT 750
     :::::|||  :::     |||  :::::::::|||:::::::::
 174 hrSerAsnLeuAsnGlnGluGlyLeuIleGlyPheGlyAspAsnSerLys 190

 751 AAACATAGCCCA.....................TATGGTTTTTTACC 776
        |||||||||                      :::  |||
 191 HisHisSerProGluLysLeuLysGluValLeuSerGlnAsnAlaLeuTh 207

 777 AACAGGAGGCTCATTTGGCGACAGTGGCTCACCAATGTTTATCTATGATG 826
      :::  :::  :::|||||||||||||||||:::|||  ||||||
 207 rAsnTyrAlaValLeuGlyAspSerGlySerProLeuPheAlaTyrAspL 224

 827 CCCAAAAGCAAAAGTGGTTAATTAATGGGGTATTGCAAACAGGCAACCCC 876
       |||:::::::::|||:::::  |||:::  :::
 224 ysGlnGluLysArgTrpValPheLeuGlyAlaTyrAspTyrTrpAlaGly 240

 877 TATATAGGAAAAAGCAATGGCTTCCAGCTAGTTCGTAAAGATTGG...TT 923
     |||      ::::::|||:::::|||            :::|||  ::
 241 Tyr......GlnLysAsnSerTrpGln...........GluTrpAsnIl 251

 924 CTATGATGAAATCTTTGCTGGAGATACCCATTCAGTATTCTACGAACCAC 973
     :|||  :::  ||||||  :::      :::
 251 eTyrLysLysGluPheAlaAspGluIleLysGln.............. 262

 974 ATCAAAATGGGAAATACTTTTTTAACGACAATAATAATGGCGCAGGAAAA 1023
                        :::||||||        ||||||
 263 ......................ArgAspAsn.........AlaGlyThr 268

1024 ATCGATGCCAAACATAAACACTATTCTCTACCTTATAGATTAAAAACACG 1073
     |||  ::::::  :::|||:::                ||||||
 269 IleLysGlyAsnGlyGluHisHis..............TrpLysThrTh 280

1074 AACCGTTCAATTGTTTAATGTTTCTTTATCCGAGACAGCAAGAGAACCTG 1123
        :::        |||      :::::::::|||
 280 rGlyThr.........AsnSerHisIleGlySerThr............. 289

1124 TTTATCATGCTGCAGGTGGGGTCAACAGTTATCGACCCAGACTGAATAAT 1173
           |||:::      :::|||:::  |||        ||||||
 290 ........AlaValArgLeuAlaAsnAsnGluArgAsp...AlaAsnAsn 302

1174 GGAGAAAATATTTCCTTTATTGACAAAGGAAAAGGTGAATTGATACTTAC 1223
     |||:::|||::::::|||      ::::::::::|||  |||:::|||
 303 GlyGlnAsnValThrPhe......GluAsnAsnGlyThrLeuValLeuAs 317

1224 CAGCAACATCAACCAAGGCGCGGGCGGTTTGTATTTTGAGGGTAATTTTA 1273
```

```
      :::|||||||||||||||||||||||||||:::|||:::|||:::::::|
  317 pGlnAsnIleAsnGlnGlyAlaGlyGlyLeuPhePheLysGlyAspTyrT 334

 1274 CGGTCTCGCCTAAAAACAACGAA...ACGTGGCAAGGCGCGGGCGTTCAT 1320
      |||||:::      |||||||:::  ||||||  ||||||||||:::
  334 hrValLysGlyIleAsnAsnAspIleThrTrpLeuGlyAlaGlyIleAsp 350

 1321 ATCAGTGATGGCAGTACCGTTACTTGGAAAGTAAACGGCGTGGCAAACGA 1370
      ::::::|||||:::   |||:::|||:::|||:::::::      :::||
  351 ValAlaAspGlyLysLysValValTrpGlnValLysAsnProAsnGlyAs 367

 1371 CCGCCTGTCCAAAATCGGCAAAGGCACGCTGCTGGTTCAAGCCAAAGGGG 1420
      |||||||:::|||||||||||||||||||||  :::::::::  |||
  367 pArgLeuAlaLysIleGlyLysGlyThrLeuGluIleAsnGlyThrGlyV 384

 1421 AAAACCAAGGCTCGGTCAGCGTGGGCGACGGTAAAGTCATCTTAGATCAG 1470
        |||||||||:::::::::|||||||||||  ||||||||||:::|||
  384 alAsnGlnGlyGlnLeuLysValGlyAspGlyThrValIleLeuAsnGln 400

 1471 CAGGCGGACGATCAAGGCAAAAAACAAGCCTTTAGTGAAATCGGCTTGGT 1520
      |||||||||  :::  |||  ||||||||||||:::::|||:::||
  401 GlnAlaAspAlaAspLysLysValGlnAlaPheSerGlnValGlyIleVa 417

 1521 CAGCGGCAGGGGGACGGTGCAACTGAATGCCGATAATCAGTTCAACCCCG 1570
      ||||||||||  |||:::  ||||||:::::|||||||:::|||||||
  417 lSerGlyArgArgThrLeuValLeuAsnSerSerAsnGlnIleAsnProA 434

 1571 ACAAACTCTATTTCGGCTTTCGCGGCGGACGTTTGGATTTGAACGGGCAT 1620
      ||:::|||||||||||||||||||||||||||||||  ||||||:::
  434 spAsnLeuTyrPheGlyPheArgGlyGlyArgLeuAspAlaAsnGlyAsn 450

 1621 TCGCTTTCGTTCCACCGCATTCAAAATACCGATGAAGGGGCGATGATTGT 1670
      :::|||:::|||:::::|||:::|||:::||||||||||||  |||||
  451 AspLeuThrPheGluHisIleArgAsnValAspGluGlyAlaArgIleVa 467

 1671 CAACCACAATCAAGACAAAGAATCCACCGTTACCATTACAGGCAATAAAG 1720
      ||||||||||            ||||||:::|||:::||||||:::::
  467 lAsnHisAsnThrGlyHisAlaSerThrIleThrLeuThrGlyLysSerL 484

 1721 ATATTACT 1728
        ||||||
  484 euIleThr 486
```

```
seq_name: pir2:S61320

seq_documentation_block:
IgA-specific metalloendopeptidase (EC 3.4.24.13) homolog - Neisseria meningitidis (fr
C;Species: Neisseria meningitidis
A;Variety: NGC65
C;Date: 20-Jul-1996 #sequence_revision 13-Mar-1997 #text_change 08-Dec-2000
C;Accession: S61320
R;Lomholt, H.; Poulsen, K.; Mogens, K.
Mol. Microbiol. 15, 495-506, 1995
A;Title: Comparative characterization of the iga gene encoding IgA1 protease in Neiss
A;Reference number: S61314; MUID:95302961
A;Accession: S61320
A;Status: preliminary; nucleic acid sequence not shown
A;Molecule type: DNA
A;Residues: 1-508 <LOM>
A;Cross-references: EMBL:X82484; NID:g732871; PIDN:CAA57867.1; PID:g732872
C;Superfamily: IgA-specific metalloendopeptidase
C;Keywords: hydrolase; metalloproteinase


alignment_scores:
            Quality:  851.50              Length:    553
              Ratio:  2.426                Gaps:     18
 Percent Similarity:  63.472   Percent Identity: 37.794

alignment_block:
 US-09-303-518D-653 x S61320   ..

 Align seg 1/1  to: S61320  from: 1  to: 508
```

```
202 GCGAAAGATATTGAGGTTTACAACAAAAAAGGGGAGTTGGTCGGCAAATC 251
    |||   |||:::::|||   ||||||:::|||:::   :::|||:::::
  1 AlaThrAspLeuSerValLysAsnLysGlnGlyGlnAsnIleGlyAsnAl 17

252 GATGACGAAAGCCCCGATGATTGATTTTTCTGTGGTATCGCGTAAC...G 298
    :::::::::::::|||||||||||||||||||||||:::::   |||
 17 aLeuSerAsnValProMetIleAspPheSerValAlaAspValAsnLysA 34

299 GCGTGGCGGCATTGGCGGGCGATCAATATATTGTGAGCGTGGCACAT... 345
      :::|||:::::::::     ||||||   ||||||||   |||
 34 rgIleAlaThrValValAspProGlnTyrAlaValSerValLysHisAla 50

346 .....:.............AACGGCGGCTATAACAATGTTGATTT 371
                        |||   ||||||:::   :::
 51 LysAlaGluValHisThrPheTyrTyrGlyGlnTyrAsnGlyHisAsnAs 67

372 TGGTGCGGAGGGAAGCAATCCCGATCAGCACCGCTTTTCTTACCAAATTG 421
       |||:::   :::|||                  :::|||:::::|
 67 pValAlaAspLysGluAsn.................GluTyrArgValV 78

422 TGAAAAGAAATAATTATAAAGCAGGGACTAACGGCCATCCTTATGGCGGC 471
    ||:::::|||||||||:::                  |||:::   :::
 78 alGluGlnAsnAsnTyrGlu.................ProHisLysAla 88

472 ..........................GATTATCATATGCCGCGTTT 491
                              ||||||:::|||   |||::
 89 TrpSerAlaSerAsnLeuGlyArgLeuGluAspTyrAsnMetAlaArgPh 105

492 GCACAAATTTGTCACAGATGCAGAACCTGTTGAGATGACCAGTTATATGG 541
    ::::|||||||||||||:::::   |||:::      |||:::
105 eAsnLysPheValThrGluValAlaProIleAlaProThrAspAlaGlyG 122

542 ATGGGTGG...AAATACGCTGATTTAAATAAATACCCTGATCGTGTTCGA 588
      |||         |||   |||   |||:::::   :::   ||||||
122 lyGlyLeuAspThrTyrLysAspLysAsnArgPheSerSerPheValArg 138

589 ATCGGAGCAGGCAGACAA......TATTGGCGGTCTGATGAAGACGAACC 632
    :::|||||||||||||||      |||   ::::::        :::
139 ValGlyAlaGlyArgGlnLeuValTyrGluLysGlyAlaTyrHisGlnGl 155

633 CAATAACCGCGAAAGTTCATAT.........CATATTGCAAGCGCATATT 673
     :::|||   |||::::::|||            :::::::::||||||
155 uGlyAsn...GluLysGlyTyrAspLeuArgAspLeuSerGlnAlaTyrA 171

674 CTTGGCTCGTCGGTGGC.............................. 690
      :::   ::::::|||
171 rgTyrAlaIleAlaGlyThrProTyrLysAspIleAsnIleAspGlnThr 187

691 ...AATACCTTTGCACAAAATGGATCAGGTGGTGGCACAGTCAACTTAGG 737
       ||||||   :::      |||   |||:::   :::::::::   ::
188 MetAsnThrGluGlyLeuIleGlyPheGlyAsnHisAsnThrHisTyrSe 204

738 TAGCGAAAAAATTAAACATAGC...CCATATGGTTTTTTACCAACAGGAG 784
    ::::|||:::::::|||:::::               |||   :::   |
204 rAlaGluGluLeuLysGlnAlaLeuSerGlnAspAlaLeuThrAsnTyrG 221

785 GCTCATTTGGCGACAGTGGCTCACCAATGTTTATCTATGATGCCCAAAAG 834
    ||   :::||||||||||||||||||:::|||   :::|||   ||||||
221 lyValLeuGlyAspSerGlySerProLeuPheAlaPheAspLysGlnLys 237

835 CAAAAGTGGTTAATTAATGGGGTATTGCAAACAGGCAACCCCTATATAGG 884
    ::::::|||::::::   |||:::   :::            |||
238 AsnGlnTrpValPheLeuGlyThrTyrAspTyrTrpAlaGlyTyrGlyLy 254

885 AAAAAGCAATGGCTTCCAGCTAGTTCGTAAAGATTGGTTCTATGATGAAA 934
     ||||||               :::   |||      |||   |||:::|
254 sLysSerTrpGlnGluTrpAsnIleTyrLysLysGluPheAlaAspLysI 271

935 TCTTTGCTGGAGATACCCATTCAGTATTCTACGAACCACATCAAAATGGG 984
    ||         |||
271 leLysGlnArgAsp.............................. 275

985 AAATACTTTTTTAACGACAATAATAATGGCGCAGGAAAAATCGATGCCAA 1034
                              |||   |||:::   |||   ::::::::
276 .....................AsnAlaGlyThrIleLysGlyAsnGlyGl 285

1035 ACATAAACACTATTCTCTACCTTATAGATTAAAAACACGAACCGTTCAAT 1084
     :|||   |||::::::::::   :::
285 uHis...HisTrpAsnIleThrPheGlyThr..................... 294

1085 TGTTTAATGTTTCTTTATCCGAGACAGCAAGAGAACCTGTTTATCATGCT 1134
          |||      :::::::::||||||                 :::
295 .....AsnSerHisIleGlySerThrAla.............ValArgLeu 305

1135 GCAGGTGGGGTCAACAGTTATCGACCCAGACTGAATAATGGAGAAAATAT 1184
     ||||||:::   ::::::              ||||||||||:::|||::
306 AlaGlyAsnGluLysAsp.............AlaAsnAsnGlyGlnAsnVa 318

1185 TTCCTTTATTGACAAAGGAAAAGGTGAATTGATACTTACCAGCAACATCA 1234
     :::::|||   |||      :::|||   |||:::|||   :::|||||||
318 lThrPheGluAsp......AsnGlyThrLeuValLeuAspGlnAsnIleA 333

1235 ACCAAGGCGCGGGCGGTTTGTATTTTGAGGGTAATTTTACGGTCTCGCCT 1284
     |||||||||||||||||||:::|||:::|||:::::::||||||:::
333 snGlnGlyAlaGlyGlyLeuPhePheLysGlyAspTyrThrValLysGly 349

1285 AAAAACAACGAA...ACGTGGCAAGGCGCGGGCGTTCATATCAGTGATGG 1331
        |||||||:::   ||||||   |||||||||:::   :::::::||||
350 AlaAsnAsnAspIleThrTrpLeuGlyAlaGlyIleAspValAlaAspGl 366

1332 CAGTACCGTTACTTGGAAAGTAAACGGCGTGGCAAACGACCGCCTGTCCA 1381
     |:::   |||:::|||:::|||::::::      :::||||||||||:::|
366 yLysLysValValTrpGlnValLysAsnProAsnGlyAspArgLeuAlaL 383

1382 AAATCGGCAAAGGCACGCTGCTGGTTCAAGCCAAAGGGGAAAACCAAGGC 1431
     ||||||||||||||||||   :::::::::   |||   |||||||||
383 ysIleGlyLysGlyThrLeuGluIleAsnGlyThrGlyValAsnGlnGly 399

1432 TCGGTCAGCGTGGGCGACGGTAAAGTCATCTTAGATCAGCAGGCGGACGA 1481
     :::::::::|||||||||||   |||||||||:::|||:::||||||::
400 GlnLeuLysValGlyAspGlyThrValIleLeuAsnGlnLysAlaAspSe 416

1482 TCAAGGCAAAAAACAAGCCTTTAGTGAAATCGGCTTGGTCAGCGGCAGGG 1531
     ::::   |||   |||||||||||:::::::|||:::||||||||||||
416 rAsnGlnLysValGlnAlaPheSerGlnValGlyIleValSerGlyArgA 433

1532 GGACGGTGCAACTGAATGCCGATAATCAGTTCAACCCCGACAAACTCTAT 1581
       |||:::   ||||||:::::|||||:::||||||||||:::||||||
433 rgThrLeuValLeuAsnSerSerAsnGlnIleAsnProAspAsnLeuTyr 449

1582 TTCGGCTTTCGCGGCGGACGTTTGGATTTGAACGGGCATTCGCTTTCGTT 1631
     |||||||||||||||||||||||||||   ||||||:::::|||:::||
450 PheGlyPheArgGlyGlyArgLeuAspAlaAsnGlyAsnAspLeuThrPh 466

1632 CCACCGCATTCAAAATACCGATGAAGGGGCGATGATTGTCAACCACAATC 1681
     |:::::::|||:::|||:::|||||||||||||   |||||||||||||||
466 eGluHisIleArgAsnValAspGluGlyAlaArgIleValAsnHisAsnT 483

1682 AAGACAAAGAATCCACCGTTACCATTACAGGCAATAAAGATATTACTACA 1731
                  ||||||:::|||:::||||||:::::   ||||||:::
483 hrGlyHisAlaSerThrIleThrLeuThrGlyLysSerLeuIleThrAla 499

1732 ACCGGCAAT 1740
          |||
500 ProGlnAsn 502
```

seq_name: pir2:S61330

seq_documentation_block:
IgA-specific metalloendopeptidase (EC 3.4.24.13) homolog - Neisseria meningitidis (f
C;Species: Neisseria meningitidis
A;Variety: SM1166
C;Date: 23-Jul-1996 #sequence_revision 13-Mar-1997 #text_change 08-Dec-2000
C;Accession: S61330
R;Lomholt, H.; Poulsen, K.; Mogens, K.

Mol. Microbiol. 15, 495-506, 1995
A;Title: Comparative characterization of the iga gene encoding IgA1 protease in Neisseri
A;Reference number: S61314; MUID:95302961
A;Accession: S61330
A;Status: preliminary; nucleic acid sequence not shown
A;Molecule type: DNA
A;Residues: 1-508 <LOM>
A;Cross-references: EMBL:X82486; NID:g732860; PIDN:CAA57869.1; PID:g732861
C;Superfamily: IgA-specific metalloendopeptidase
C;Keywords: hydrolase; metalloproteinase

alignment_scores:
Quality: 845.50 Length: 549
Ratio: 2.416 Gaps: 18
Percent Similarity: 63.752 Percent Identity: 37.523

alignment_block:
US-09-303-518D-653 x S61330 ..

Align seg 1/1 to: S61330 from: 1 to: 508

```
202 GCGAAAGATATTGAGGTTTACAACAAAAAAGGGGAGTTGGTCGGCAAATC 251
    |||   |||:::::|||   |||||||:::|||:::   :::|||:::::
  1 AlaThrAspLeuSerValLysAsnLysGlnGlyGlnAsnIleGlyAsnAl 17

252 GATGACGAAAGCCCCGATGATTGATTTTTCTGTGGTATCGCGTAAC...G 298
    :::::::::::::|||||||||||||||||||||||:::::   |||
 17 aLeuSerAsnValProMetIleAspPheSerValAlaAspValAsnLysA 34

299 GCGTGGCGGCATTGGCGGGCGATCAATATATTGTGAGCGTGGCACAT... 345
      :::|||:::::::::      ||||||   |||||||||   |||
 34 rgIleAlaThrValValAspProGlnTyrAlaValSerValLysHisAla 50

346 .......................AACGGCGGCTATAACAATGTTGATTT 371
                            |||   |||||||:::   :::
 51 LysAlaGluValHisThrPheTyrTyrGlyGlnTyrAsnGlyHisAsnAs 67

372 TGGTGCGGAGGGAAGCAATCCCGATCAGCACCGCTTTTCTTACCAAATTG 421
      |||:::   :::|||                  :::|||:::::::|
 67 pValAlaAspLysGluAsn.................GluTyrArgValV 78

422 TGAAAAGAAATAATTATAAAGCAGGGACTAACGGCCATCCTTATGGCGGC 471
    ||::::::|||||||||:::                  |||:::   :::
 78 alGluGlnAsnAsnTyrGlu.................ProHisLysAla 88

472 ..............................GATTATCATATGCCGCGTTT 491
                                  |||||||:::|||   |||::
 89 TrpSerAlaSerAsnLeuGlyArgLeuGluAspTyrAsnMetAlaArgPh 105

492 GCACAAATTTGTCACAGATGCAGAACCTGTTGAGATGACCAGTTATATGG 541
    ::::|||||||||||||:::::   |||:::   |||:::
105 eAsnLysPheValThrGluValAlaProIleAlaProThrAspAlaGlyG 122

542 ATGGGTGG...AAATACGCTGATTTAAATAAATACCCTGATCGTGTTCGA 588
      |||         |||   |||   |||::::::   :::   ||||||
122 lyGlyLeuAspThrTyrLysAspLysAsnArgPheSerSerPheValArg 138

589 ATCGGAGCAGGCAGACAA......TATTGGCGGTCTGATGAAGACGAACC 632
    :::|||||||||||||||      |||   ::::::          :::
139 ValGlyAlaGlyArgGlnLeuValTyrGluLysGlyAlaTyrHisGlnGl 155

633 CAATAACCGCGAAAGTTCATAT.........CATATTGCAAGCGCATATT 673
     :::|||   |||:::::::|||            :::::::::||||||
155 uGlyAsn...GluLysGlyTyrAspLeuArgAspLeuSerGlnAlaTyrA 171

674 CTTGGCTCGTCGGTGGC................................ 690
      :::   :::::::|||
171 rgTyrAlaIleAlaGlyThrProTyrLysAspIleAsnIleAspGlnThr 187

691 ...AATACCTTTGCACAAAATGGATCAGGTGGTGGCACAGTCAACTTAGG 737
      ||||||   :::     |||   |||:::   :::::::::   ::
188 MetAsnThrGluGlyLeuIleGlyPheGlyAsnHisAsnThrHisTyrSe 204

738 TAGCGAAAAAATTAAACATAGC...CCATATGGTTTTTTACCAACAGGAG 784
    :::::|||:::::::|||::::::                |||   :::   |
204 rAlaGluGluLeuLysGlnAlaLeuSerGlnAspAlaLeuThrAsnTyrG 221

785 GCTCATTTGGCGACAGTGGCTCACCAATGTTTATCTATGATGCCCAAAAG 834
    ||   :::|||||||||||||||||||:::|||   :::|||   ||||||
221 lyValLeuGlyAspSerGlySerProLeuPheAlaPheAspLysGlnLys 237

835 CAAAAGTGGTTAATTAATGGGGTATTGCAAACAGGCAACCCCTATATAGG 884
    :::::::|||::::::   |||:::   :::              |||
238 AsnGlnTrpValPheLeuGlyThrTyrAspTyrTrpAlaGlyTyrGlyLy 254

885 AAAAAGCAATGGCTTCCAGCTAGTTCGTAAAGATTGGTTCTATGATGAAA 934
     ||||||                :::   |||       |||   |||:::|
254 sLysSerTrpGlnGluTrpAsnIleTyrLysLysGluPheAlaAspLysI 271

935 TCTTTGCTGGAGATACCCATTCAGTATTCTACGAACCACATCAAAATGGG 984
    ||         |||
271 leLysGlnArgAsp..................................... 275

985 AAATACTTTTTTAACGACAATAATAATGGCGCAGGAAAAATCGATGCCAA 1034
                              |||   |||:::   |||   ::::::::
276 .....................AsnAlaGlyThrIleLysGlyAsnGlyGl 285

1035 ACATAAACACTATTCTCTACCTTATAGATTAAAAACACGAACCGTTCAAT 1084
    :|||   |||::::::::::   :::
285 uHis...HisTrpAsnIleThrPheGlyThr..................... 294

1085 TGTTTAATGTTTCTTTATCCGAGACAGCAAGAGAACCTGTTTATCATGCT 1134
       |||       :::::::::||||||                 :::
295 .....AsnSerHisIleGlySerThrAla.............ValArgLeu 305

1135 GCAGGTGGGGTCAACAGTTATCGACCCAGACTGAATAATGGAGAAAATAT 1184
     |||||||:::   ::::::              ||||||||||:::|||::
306 AlaGlyAsnGluArgAsp.............AlaAsnAsnGlyGlnAsnVa 318

1185 TTCCTTTATTGACAAAGGAAAAGGTGAATTGATACTTACCAGCAACATCA 1234
     :::::|||   |||       :::|||   |||:::|||   :::|||||||
318 lThrPheGluAsp......AsnGlyThrLeuValLeuAspGlnAsnIleA 333

1235 ACCAAGGCGCGGGCGGTTTGTATTTTGAGGGTAATTTTACGGTCTCGCCT 1284
     |||||||||||||||||||:::|||:::|||::::::|||||:::
333 snGlnGlyAlaGlyLeuPhePheLysGlyAspTyrThrValLysGly 349

1285 AAAAACAACGAA...ACGTGGCAAGGCGCGGGCGTTCATATCAGTGATGG 1331
       ||||||:::   ||||||   ||||||||||:::   ::::::|||||
350 AlaAsnAsnAspIleThrTrpLeuGlyAlaGlyIleAspValAlaAspGl 366

1332 CAGTACCGTTACTTGGAAAGTAAACGGCGTGGCAAACGACCGCCTGTCCA 1381
     |:::   |||:::|||:::|||::::::       :::|||||||||:::|
366 yLysLysValValTrpGlnValLysAsnProAsnGlyAspArgLeuAlaL 383

1382 AAATCGGCAAAGGCACGCTGCTGGTTCAAGCCAAAGGGGAAAACCAAGGC 1431
     ||||||||||||||||||||   :::::::::   |||   |||||||||
383 ysIleGlyLysGlyThrLeuGluIleAsnGlyThrGlyValAsnGlnGly 399

1432 TCGGTCAGCGTGGGCGACGGTAAAGTCATCTTAGATCAGCAGGCGGACGA 1481
     :::::::::|||||||||||||   |||||||||:::|||:::|||||::
400 GluLeuLysValGlyAspGlyThrValIleLeuAsnGlnLysAlaAspSe 416

1482 TCAAGGCAAAAAACAAGCCTTTAGTGAAATCGGCTTGGTCAGCGGCAGGG 1531
     ::::   |||   :::|||||||||:::::::|||:::|||   |||||||
416 rAsnGlnLysValSerAlaPheSerGlnValGlyIleValArgGlyArgG 433

1532 GGACGGTGCAACTGAATGCCGATAATCAGTTCAACCCCGACAAACTCTAT 1581
     |||||:::   |||||||:::   :::|||:::|||||||:::::|||||
433 lyThrLeuValLeuAsnSerProAspGlnIleAsnProAsnAsnLeuTyr 449

1582 TTCGGCTTTCGCGGCGGACGTTTGGATTTGAACGGGCATTCGCTTTCGTT 1631
     |||||||||||||||||||||||||||   ||||||:::::|||:::||
450 PheGlyPheArgGlyGlyArgLeuAspAlaAsnGlyAsnAspLeuThrPh 466
```

```
  1632 CCACCGCATTCAAAATACCGATGAAGGGGCGATGATTGTCAACCACAATC 1681
       |:::::::|||:::|||:::|||||||||||||   |||||||||||||||
   466 eGluHisIleArgAsnValAspGluGlyAlaArgIleValAsnHisAsnT 483

  1682 AAGACAAAGAATCCACCGTTACCATTACAGGCAATAAAGATATTACT 1728
         |||:::   ||||||:::|||:::|||||:::::   ||||||
   483 hrAspArgAlaSerThrIleThrLeuThrGlyLysSerLeuIleThr 498

seq_name: pir2:S61331

seq_documentation_block:
IgA-specific metalloendopeptidase (EC 3.4.24.13) homolog - Neisseria meningitidis (fragm
C;Species: Neisseria meningitidis
A;Variety: NGO93
C;Date: 23-Jul-1996 #sequence_revision 13-Mar-1997 #text_change 08-Dec-2000
C;Accession: S61331
R;Lomholt, H.; Poulsen, K.; Mogens, K.
Mol. Microbiol. 15, 495-506, 1995
A;Title: Comparative characterization of the iga gene encoding IgA1 protease in Neisseri
A;Reference number: S61314; MUID:95302961
A;Accession: S61331
A;Status: preliminary; nucleic acid sequence not shown
A;Molecule type: DNA
A;Residues: 1-507 <LOM>
A;Cross-references: EMBL:X82482; NID:g732885; PIDN:CAA57865.1; PID:g732886
C;Superfamily: IgA-specific metalloendopeptidase
C;Keywords: hydrolase; metalloproteinase


alignment_scores:
            Quality:  840.00              Length:    546
              Ratio:  2.421                Gaps:     17
 Percent Similarity:  63.553   Percent Identity: 37.912

alignment_block:
 US-09-303-518D-653 x S61331   ..

 Align seg 1/1  to: S61331  from: 1  to: 507

    202 GCGAAAGATATTGAGGTTTACAACAAAAAAGGGGAGTTGGTCGGCAAATC 251
        |||   |||:::::::|||   ||||||:::|||:::   :::|||:::::
      1 AlaThrAspLeuSerValLysAsnLysGlnGlyGlnAsnIleGlyAsnAl 17

    252 GATGACGAAAGCCCCGATGATTGATTTTTCTGTGGTATCGCGTAAC...G 298
        :::::::::::::|||||||||||||||||||||||:::::::   |||
     17 aLeuSerAsnValProMetIleAspPheSerValAlaAspValAsnLysA 34

    299 GCGTGGCGGCATTGGCGGGCGATCAATATATTGTGAGCGTGGCACAT... 345
          :::|||::::::::::      ||||||   |||||||||   |||
     34 rgIleAlaThrValValAspProGlnTyrAlaValSerValLysHisAla 50

    346 .......................AACGGCGGCTATAACAATGTTGATTT 371
                                  |||   |||||||:::   :::
     51 LysAlaGluValHisThrPheTyrTyrGlyGlnTyrAsnGlyHisAsnAs 67

    372 TGGTGCGGAGGGAAGCAATCCCGATCAGCACCGCTTTTCTTACCAAATTG 421
          |||:::   :::|||                  :::|||::::::|
     67 pValAlaAspLysGluAsn.................GluTyrArgValV 78

    422 TGAAAAGAAATAATTATAAAGCAGGGACTAACGGCCATCCTTATGGCGGC 471
        ||:::::::||||||||||:::               |||       :|||
     78 alGluGlnAsnAsnTyrGluPro............HisLysAlaTrpGly 90

    472 .......................GATTATCATATGCCGCGTTTGCACAA 497
                                 ||||||:::|||   |||:::::||
     91 AlaSerAsnLeuGlyArgLeuGluAspTyrAsnMetAlaArgPheAsnLy 107

    498 ATTTGTCACAGATGCAGAACCTGTTGAGATGACCAGTTATATGGATGGGT 547
        ||||||||||:::::   |||:::      |||:::         |||
    107 sPheValThrGluValAlaProIleAlaProThrAspAlaGlyGlyGlyL 124

    548 GG...AAATACGCTGATTTAAATAAATACCCTGATCGTGTTCGAATCGGA 594
               |||   |||   |||::::::   :::   ||||||:::|||

    124 euAspThrTyrLysAspLysAsnArgPheSerSerPheValArgValGly 140

    595 GCAGGCAGACAA...TATTGGCGGTCTGATGAAGACGAACCCAATAACCG 641
        ||||||||||||      :::::::::      :::   ::::::::
    141 AlaGlyArgGlnLeuValTyrGluLysGlyAlaTyrHisGlnGluGlyLy 157

    642 CGAAAGTTCATAT.........CATATTGCAAGCGCATATTCTTGGCTCG 682
        :|||:::::::|||         ::::::::::||||||   :::   :
    157 sGluLysGlyTyrAspLeuArgAspLeuSerGlnAlaTyrArgTyrAlaI 174

    683 TCGGTGGC.................................AATACC 696
        ::::::|||                                 ||||||
    174 leAlaGlyThrProTyrLysAspIleAsnIleAspGlnThrMetAsnThr 190

    697 TTTGCACAAAATGGATCAGGTGGTGGCACAGTCAACTTAGGTAGCGAAAA 746
          :::     |||   |||:::   :::   :::   ::::::|||::
    191 GluGlyLeuIleGlyPheGlyAsnHisAsnLysGlnTyrSerAlaGluGl 207

    747 AATTAAACATAGC...CCATATGGTTTTTTACCAACAGGAGGCTCATTTG 793
        ::::|||::::::          |||   :::   |||   :::|
    207 uLeuLysGlnAlaLeuSerGlnAspAlaLeuThrAsnTyrGlyValLeuG 224

    794 GCGACAGTGGCTCACCAATGTTTATCTATGATGCCCAAAAGCAAAAGTGG 843
        |||||||||||||||||:::|||   :::|||   |||||::::::|||
    224 lyAspSerGlySerProLeuPheAlaPheAspLysGlnLysAsnGlnTrp 240

    844 TTAATTAATGGGGTATTGCAAACAGGCAACCCCTATATAGGAAAAAGCAA 893
        ::::::   |||:::   :::          |||      ||||||
    241 ValPheLeuGlyThrTyrAspTyrTrpAlaGlyTyrGlyLysLysSerTr 257

    894 TGGCTTCCAGCTAGTTCGTAAAGATTGGTTCTATGATGAAATCTTTGCTG 943
                  :::   |||      |||   |||||||||
    257 pGlnGluTrpAsnIleTyrLysLysGluPheAlaAspGluIle....... 271

    944 GAGATACCCATTCAGTATTCTACGAACCACATCAAAATGGGAAATACTTT 993

    271 ..................................................... 271

    994 TTTAACGACAATAATAATGGCGCAGGAAAAATCGATGCCAAACATAAACA 1043
              :::::::::::::||||||   |||   :::       :::||
    272 ......LysGlnArgAspAsnAlaGlyThrIleLysGlyTyrGlyGluHi 286

   1044 CTATTCTCTACCTTATAGATTAAAAACACGAACCGTTCAATTGTTTAATG 1093
        |:::                  ||||||      :::          |||
    286 sHis..............TrpLysThrThrGlyThr.........AsnS 295

   1094 TTTCTTTATCCGAGACAGCAAGAGAACCTGTTTATCATGCTGCAGGTGGG 1143
             :::::::::|||                        |||:::
    295 erHisIleGlySerThr.....................AlaValArgLeu 304

   1144 GTCAACAGTTATCGACCCAGACTGAATAATGGAGAAAATATTTCCTTTAT 1193
        :::|||:::   |||        |||||||||:::|||::::::|||
    305 AlaAsnAsnGluArgAsp...AlaAsnAsnGlyGlnAsnValThrPhe.. 319

   1194 TGACAAAGGAAAAGGTGAATTGATACTTACCAGCAACATCAACCAAGGCG 1243
             :::::::::|||   |||:::||| ' :::|||||||||||||||||
    320 ....GluAsnAsnGlyThrLeuValLeuAspGlnAsnIleAsnGlnGlyA 335

   1244 CGGGCGGTTTGTATTTTGAGGGTAATTTTACGGTCTCGCCTAAAAACAAC 1293
        ||||||||||||:::|||:::|||::::::||||||:::      ||||||
    335 laGlyGlyLeuPhePheLysGlyAspTyrThrValLysGlyAlaAsnAsn 351

   1294 GAA...ACGTGGCAAGGCGCGGGCGTTCATATCAGTGATGGCAGTACCGT 1340
        :::   ||||||   |||||||||:::   ::::::||||||:::   ||
    352 AspIleThrTrpLeuGlyAlaGlyIleAspValAlaAspGlyLysLysVa 368

   1341 TACTTGGAAAGTAAACGGCGTGGCAAACGACCGCCTGTCCAAAATCGGCA 1390
        |:::|||:::|||::::::      :::||||||||||:::||||||||||
    368 lValTrpGlnValLysAsnProAsnGlyAspArgLeuAlaLysIleGlyL 385

   1391 AAGGCACGCTGCTGGTTCAAGCCAAAGGGGAAAACCAAGGCTCGGTCAGC 1440
        ||||||||||   :::::::::   |||   |||||||||:::::::::
    385 ysGlyThrLeuGluIleAsnGlyThrGlyValAsnGlnGlyGlnLeuLys 401
```

```
1441 GTGGGCGACGGTAAAGTCATCTTAGATCAGCAGGCGGACGATCAAGGCAA 1490
     ||||||||||||   |||||||||:::|||:::||||||   :::   ||
 402 ValGlyAspGlyThrValIleLeuAsnGlnLysAlaAspAlaAspLysLy 418

1491 AAAACAAGCCTTTAGTGAAATCGGCTTGGTCAGCGGCAGGGGGACGGTGC 1540
     |   ||||||||||||:::::|||:::||||||||||||   |||:::
 418 sValGlnAlaPheSerGlnValGlyIleValSerGlyArgArgThrLeuV 435

1541 AACTGAATGCCGATAATCAGTTCAACCCCGACAAACTCTATTTCGGCTTT 1590
       ||||||:::::|||||:::|||||||||:::|||||||||||||||
 435 alLeuAsnSerSerAsnGlnIleAsnProAspAsnLeuTyrPheGlyPhe 451

1591 CGCGGCGGACGTTTGGATTTGAACGGGCATTCGCTTTCGTTCCACCGCAT 1640
     ||||||||||||||||   ||||||:::::|||:::|||:::::||
 452 ArgGlyGlyArgLeuAspAlaAsnGlyAsnAspLeuThrPheGluHisIl 468

1641 TCAAAATACCGATGAAGGGGCGATGATTGTCAACCACAATCAAGACAAAG 1690
     |:::|||:::||||||||||   |||||||||||||||
 468 eArgAsnValAspGluGlyAlaArgIleValAsnHisAsnThrGlyHisA 485

1691 AATCCACCGTTACCATTACAGGCAATAAAGATATTACT 1728
       ||||||:::|||:::||||||:::::   ||||||
 485 laSerThrIleThrLeuThrGlyLysSerLeuIleThr 497
```

seq_name: pir2:S61332

seq_documentation_block:
IgA-specific metalloendopeptidase (EC 3.4.24.13) homolog - Haemophilus influenzae (strai
N;Alternate names: IgA1 protease
C;Species: Haemophilus influenzae
A;Variety: HK635
C;Date: 14-Feb-1997 #sequence_revision 13-Mar-1997 #text_change 08-Dec-2000
C;Accession: S61332
R;Lomholt, H.; Poulsen, K.; Mogens, K.
Mol. Microbiol. 15, 495-506, 1995
A;Title: Comparative characterization of the iga gene encoding IgA1 protease in Neisseri
A;Reference number: S61314; MUID:95302961
A;Accession: S61332
A;Status: preliminary; nucleic acid sequence not shown
A;Molecule type: DNA
A;Residues: 1-462 <LOM>
A;Cross-references: EMBL:X82488; NID:g732714; PIDN:CAA57871.1; PID:g732715
A;Experimental source: strain HK635
A;Note: the authors did not translate the codon for residue 462
C;Genetics:
A;Gene: iga
C;Superfamily: IgA-specific metalloendopeptidase
C;Keywords: hydrolase; metalloproteinase

alignment_scores:
Quality: 718.50 Length: 499
Ratio: 2.267 Gaps: 17
Percent Similarity: 63.527 Percent Identity: 36.072

alignment_block:
US-09-303-518D-653 x S61332 ..

Align seg 1/1 to: S61332 from: 1 to: 462

```
 202 GCGAAAGATATTGAGGTTTACAACAAAAAAGGGGAGTTGGTCGGCAAATC 251
     |||   ::::::||||||   :::|||:::::::::   :::|||:::::
   1 AlaThrAsnValGluValArgAspLysAsnAsnHisSerLeuGlyAsnAl 17

 252 GATGACGAAAGCC...CCGATGATTGATTTTTCTGTGGTATCGCGTAAC. 297
     ::::   ::::::   |||||||||||||||||||||||:::   |||
  17 aLeuProAsnGlyIleProMetIleAspPheSerValValAspValAsnL 34

 298 ..GGCGTGGCGGCATTGGCGGGCGATCAATATATTGTGAGCGTGGCACAT 345
         ::::::::::|||::      ||||||||||||||||||   |||
  34 ysArgIleGlyThrLeuValAspProGlnTyrIleValSerValLysHis 50

 346 AACGGCGGCTATAACAATGTTGATTTT.................GGTGC 377
             |||   |||   ||||||                    |||
  51 AlaHisGlnTyrMetAsn...AspPheTyrPheGlyHisTyrAsnGlyHi 66

 378 GGAGGGAAGCAATCCCGATCAGCACCGCTTTTCTTACCAAATTGTGAAAA 427
      :::      :::   |||:::::::::      |||:::::|||   :
  66 sArgAspValSerAspAspGluAsnLys......TyrSerValValThrG 81

 428 GAAATAATTATAAAGCAGGGACTAACGGCCAT......CCTTATGGCGGC 471
     ::|||||   :::   :::   |||   |||
  81 lnAsnAsnValAsnProAsnGluAsnTrpHisValAspLysArgLeuAsp 97

 472 GATTATCATATGCCGCGTTTGCACAAATTTGTCACAGATGCAGAACCTGT 521
     ||||||:::|||||||||||:::||||||||||||:::::   |||::
  98 AspTyrAsnMetProArgLeuAsnLysPheValThrGluValAlaProTh 114

 522 TGAGATGACCAGTTATATGGATGGGTGGAAA...TACGCTGATTTAAATA 568
     :      |||         |||      :::   |||   |||   :::|
 114 rThrProThrLeuAlaGlyAspAspLeuGluThrTyrLysAspLysGluL 131

 569 AATACCCTGATCGTGTTCGAATCGGAGCAGGCAGACAA......TATTGG 612
     |||||   :::   ||||||:::|||||||||||||||      |||
 131 ysTyrLeuSerPheValArgValGlyAlaGlyArgGlnLeuValTyrGlu 147

 613 CGGTCTGATGAA......GACGAACCCAATAACCGCGAAAGTTCATATCA 656
     ::::::::::::      ::::::   ::::::   |||:::
 148 LysGlySerHisHisValGluAspLysGluHisGlyGluAspLeuLysAs 164

 657 TATTGCAAGCGCATATTCTTGGCTCGTCGGTGGCAATACCTTTGCA.... 702
      :::::::::||||||   :::   :::||||||:::   :::
 164 pLeuSerAlaAlaTyrArgTyrAlaIleGlyGlyThrProTyrLysGlyI 181

 703 ..........CAAAATGGATCAGGTGGTGGCACAGTCAACTTAGGTAGC 741
                 :::   |||      |||   ::::::::::|||:::
 181 leAsnIleAspProSerGlnSerLysLysGlyLeuIleGlyPheGlyAsp 197

 742 GAAAAAATTAAACAT................................. 756
     ::::::      |||
 198 SerArgGluAspHisValIleAsnSerLysThrLeuLeuSerGlnAspPr 214

 757 ....AGCCCATATGGTTTTTTACCAACAGGAGGCTCATTTGGCGACAGTG 802
         :::   ||||||   |||                  ||||||||||
 214 oLeuThrAsnTyrGlyValLeu.................GlyAspSerG 225

 803 GCTCACCAATGTTTATCTATGATGCCCAAAAGCAAAAGTGGTTAATTAAT 852
     |||||||:::|||   :::|||   |||::::::||||||::::::
 225 lySerProLeuPheAlaPheAspLysGlnGlnAsnLysTrpValPheIle 241

 853 GGGGTATTGCAAACAGGCAACCCCTATATAGGAAAAAGCAATGGCTTCCA 902
     |||                  |||      ||||||
 242 GlyProTyrThrTyrTrpAlaGlyTyrGlyLysLysSerTrpGlnGluTr 258

 903 GCTAGTTCGTAAAGATTGGTTCTATGATGAAATCTTTGCTGGAGATACCC 952
         :::   |||:::   |||      :::::::::      |||:::
 258 pAsnIleTyrLysSerGlnPheThrLysAspValLeuAsnLysAspSerA 275

 953 ATTCAGTATTCTACGAACCACATCAAAATGGGAAATACTTTTTTAACGAC 1002
       :::::::::              |||   :::|||   :::::::::
 275 laGlyLeuLeu.........LysGlyAsnThrGlnTyrAsnTrpThrSer 288

1003 AATAATAATGGCGCAGGAAAAATCGATGCCAAACATAAACACTATTCTCT 1052
     |||:::|||   :::
 289 AsnGlyAsnThrSer.................................. 293

1053 ACCTTATAGATTAAAAACACGAACCGTTCAATTGTTTAATGTTTCTTTAT 1102
               :::   ::::::      :::|||:::::::|||:::|||
 294 .........MetIleSerAsnGlySerGluLeuLeuGluValAsnLeuP 307

1103 CCGAGACAGCAAGAGAACCTGTTTATCATGCTGCAGGTGGGGTCAACAGT 1152
       ::::::::::::                                ::::::
 307 heAspAsnSerLys.............................HisThr 313

1153 TATCGACCCAGACTGAATAATGGAGAAAATATTTCCTTTATTGACAAAGG 1202
```

```
          |||   :::   |||   |||:::::::::::::|||      :::||
      314 AsnArgGluLysAlaAsnTyrGlyLysSerValThrPhe......GlnGl 328

     1203 AAAAGGTGAATTGATACTTACCAGCAACATCAACCAAGGCGCGGGCGGTT 1252
          |:::|||   |||   |||   ::::::||||||||||||||||||||||
      328 yAsnGlyThrLeuThrLeuLysAsnSerIleAsnGlnGlyAlaGlyGlyL 345

     1253 TGTATTTTGAGGGTAATTTTACGGTC...TCGCCTAAAAACAACGAAACG 1299
          ||:::|||||||||||||:::||||||   :::   :::::::|||   :::
      345 euPhePheGluGlyAsnTyrThrValGluGlySerSerAspAsnIleVal 361

     1300 TGGCAAGGCGCGGGCGTTCATATCAGTGATGGCAGTACCGTTACTTGGAA 1349
          |||:::||||||||||:::   :::::::::|||:::|||||||||||||||
      362 TrpAsnGlyAlaGlyIleSerValAlaGluGlyLysThrValThrTrpLy 378

     1350 AGTAAACGGCGTGGCAAACGACCGCCTGTCCAAAATCGGCAAAGGCACGC 1399
          ||||::::::      :::||||||||:::|||||||||||||||||||
      378 sValHisAsnProGlnSerAspArgLeuAlaLysIleGlyLysGlyThrL 395

     1400 TGCTGGTTCAAGCCAAAGGGGAAAACCAAGGCTCGGTCAGCGTGGGCGAC 1449
          ||:::|||:::::|||||||||||||:::||||||:::::|||||||||
      395 euIleValGluGlyLysGlyGluAsnLysGlySerLeuLysValGlyAsp 411

     1450 GGTAAAGTCATCTTAGATCAGCAGGCGGACGATCAAGGCAAAAAACAAGC 1499
          |||   |||||||||   ||||||||||||   ::::::|||   :::||
      412 GlyThrValIleLeuLysGlnGlnAlaAspAlaAsnAsnLysValLysAl 428

     1500 CTTTAGTGAAATCGGCTTGGTCAGCGGCAGGGGGACGGTGCAACTGAATG 1549
          |||||||:::::|||:::||||||||||||:::||||||   ||||||
      428 aPheSerGlnValGlyIleValSerGlyArgSerThrValValLeuAsnA 445

     1550 CCGATAATCAGTTCAACCCCGACAAACTCTATTTCGGCTTTCGCGGC 1596
            |||:::|||   :::|||:::::::::|||||||||||||||||||
      445 spAspLysGlnValAspProAsnSerIleTyrPheGlyPheArgGly 460

seq_name: pir2:S61333

seq_documentation_block:
IgA-specific metalloendopeptidase (EC 3.4.24.13) homolog - Haemophilus influenzae (strai
N;Alternate names: IgA1 protease
C;Species: Haemophilus influenzae
A;Variety: HK284
C;Date: 14-Feb-1997 #sequence_revision 13-Mar-1997 #text_change 08-Dec-2000
C;Accession: S61333
R;Lomholt, H.; Poulsen, K.; Mogens, K.
Mol. Microbiol. 15, 495-506, 1995
A;Title: Comparative characterization of the iga gene encoding IgA1 protease in Neisseri
A;Reference number: S61314; MUID:95302961
A;Accession: S61333
A;Status: preliminary; nucleic acid sequence not shown
A;Molecule type: DNA
A;Residues: 1-471 <LOM>
A;Cross-references: EMBL:X82487; NID:g773205; PIDN:CAA57870.1; PID:g773206
A;Experimental source: strain HK284
C;Genetics:
A;Gene: iga
C;Superfamily: IgA-specific metalloendopeptidase
C;Keywords: hydrolase; metalloproteinase


alignment_scores:
            Quality:  695.50            Length:    514
              Ratio:  2.153               Gaps:     22
 Percent Similarity:  62.840   Percent Identity: 34.825

alignment_block:
 US-09-303-518D-653 x S61333   ..

 Align seg 1/1  to: S61333  from: 1  to: 471

      202 GCGAAAGATATTGAGGTTTACAACAAAAAAGGGGAGTTGGTCGGCAAATC 251
          |||   ::::::||||||   :::|||:::::::::   :::|||:::::
        1 AlaThrAsnValGluValArgAspLysAsnAsnGlnSerLeuGlySerAl 17

      252 GATGACGAAAGCC...CCGATGATTGATTTTTCTGTGGTATCGCGTAAC. 297
          ::::   ::::::   |||||||||||||||||||||||:::   |||
       17 aLeuProAsnGlyIleProMetIleAspPheSerValValAspValAsnL 34

      298 ..GGCGTGGCGGCATTGGCGGGCGATCAATATATTGTGAGCGTGGCACAT 345
              ::::::::::|||:::      ||||||||||||||||||   |||
       34 ysArgIleGlyThrLeuValAspProGlnTyrIleValSerValLysHis 50

      346 AAC.........................GGCGGCTATAACAATGTTGA 368
                                      |||   ||||||
       51 AlaHisLysHisIleAsnAspPheTyrPheGlyHisTyrAsn........ 64

      369 TTTTGGTGCGGAGGGAAGCAATCCCGATCAGCACCGCTTTTCTTACCAAA 418
              |||   :::       :::   |||:::::::::      |||::::
       65 ....GlyHisArgAspValSerAspAspGluAsnLys......TyrSerV 78

      419 TTGTGAAAAGAAATAATTATAAAGCAGGGACTAACGGCCAT......CCT 462
          ::|||   :::||||||   |||   :::   :::   |||
       78 alValThrGlnAsnAsnValLysProAsnGluAspTrpHisValAspLys 94

      463 TATGGCGGCGATTATCATATGCCGCGTTTGCACAAATTTGTCACAGATGC 512
                    ||||||:::||||||||||||:::|||||||||||||:::::
       95 ArgLeuAspAspTyrAsnMetProArgLeuAsnLysPheValThrGluVa 111

      513 AGAACCTGTTGAGATGACCAGTTATATGGATGGGTGGAAA...TACGCTG 559
          :   |||:::      |||         |||      :::   |||   |
      111 lAlaProThrThrProThrLeuAlaGlyAspAspLeuGluThrTyrLysA 128

      560 ATTTAAATAAATACCCTGATCGTGTTCGAATCGGAGCAGGCAGACAATAT 609
          ||   :::||||||||||:::   ||||||:::|||||||||   |||:::
      128 spLysGluLysTyrProSerPheValArgValGlyAlaGlyThrGlnPhe 144

      610 TGGCGGTCTGAT..............GAAGACGAACCCAATAACCGCGA 644
                 :::                 |||      :::|||:::::
      145 ValTyrGluLysGlySerTyrTyrValGluLysThrThrArgAsnAsnAs 161

      645 AAGTTCATATCATATTGCAAGCGCATATTCTTGGCTCGTCGGTGGCAATA 694
          :   ::::::   :::   :::||||||   :::   :::|||||||:::
      161 pIleLysPhe...LeuAspGluAlaTyrArgTyrAlaIleGlyGlyThrP 177

      695 CCTTTGCA..............CAAAATGGATCAGGTGGTGGCACAGTC 729
                 :::                 |||      |||   :::
      177 roTyrGluGlyIleAsnIleAspProSerGlnSerLysLysGlyLeuIle 193

      730 AACTTAGGTAGCGAAAAAATTAAACAT..................... 756
          ::::::|||:::::::::   :::|||
      194 GlyPheGlyAspSerArgGluAsnHisValIleAspAlaLysThrLeuLe 210

      757 ..............AGCCCATATGGTTTTTTACCAACAGGAGGCTCAT 790
                        :::   ||||||   |||
      210 uSerGlnAspProLeuThrAsnTyrGlyValLeu.............. 221

      791 TTGGCGACAGTGGCTCACCAATGTTTATCTATGATGCCCAAAAGCAAAAG 840
          ||||||||||||||||||:::|||   :::|||   |||::::::|||
      222 ..GlyAspSerGlySerProLeuPheAlaPheAspLysGlnGlnAsnLys 237

      841 TGGTTAATTAATGGGGTATTGCAAACAGGCAACCCCTAT.......... 879
          |||::::::   |||                |||||
      238 TrpValPheIleGly.................ProTyrThrTyrTrpAl 248

      880 ....ATAGGAAAAAGCAAT......GGCTTCCAGCTAGTTCGTAAAGATT 919
              ||||||::::::         :::::::::   :::||||||
      248 aGlyTyrGlyLysLysSerTrpGlnGluTrpAsnIleTyrLysLysAsp. 264

      920 GGTTCTATGATGAAATCTTTGCTGGAGATACCCATTCAGTATTCTACGAA 969
          |||   |||:::|||         |||:::   ::::::
      265 ..PheAlaAspAsnIleLysLysArgAspAsnAlaGluAlaVal...... 278

      970 CCACATCAAAATGGGAAATACTTTTTTAACGACAATAATAATGGCGCAGG 1019
          |||   :::::::::::|||   :::::::::::::::|||      ::
      279 ProPheSerThrSerGluTyrHisTrpThrAsnThrThrAsnHisGlnSe 295
```

```
 1020 AAAAATCGATGCCAAACATAAACACTATTCTCTACCTTATAGATTAAAAA 1069
      :::|||      |||:::      :::::::::          |||   :
  295 rGluIle......LysAsnThrAspHisThrIleThrValThrLeuProS 310

 1070 CACGAACCGTTCAATTGTTTAATGTTTCTTTATCCGAGACAGCAAGAGAA 1119
      ::          :::|||   |||   :::   :::|||
  310 erAspProAsnArgLeuValAsnPheGlnGlnLysGlu............ 322

 1120 CCTGTTTATCATGCTGCAGGTGGGGTCAACAGTTATCGACCCAGACTGAA 1169
                                          :::|||::
  323 .......................................HisLeuGl 325

 1170 TAATGGAGAAAATATTTCCTTTATTGAC.........AAAGGAAAAGGTG 1210
      :::|||:::|||::::::|||   |||         :::|||||||||
  325 nThrGlyGlnAsnValThrPheAspAspSerThrAsnAsnGlyLysGlyT 342

 1211 AATTGATACTTACCAGCAACATCAACCAAGGCGCGGGCGGTTTGTATTTT 1260
        |||||||||   ::::::|||||||||||||||||||||||||:::|||
  342 hrLeuIleLeuAspAspHisIleAsnGlnGlyAlaGlyGlyLeuPhePhe 358

 1261 GAGGGTAATTTTACGGTCTCGCCTAAAAACAACGAA...ACGTGGCAAGG 1307
      :::|||||:::   |||:::   |||:::::::::   ||||||   ||
  359 LysGlyAsnTyrGluValLysGlyLysThrAspAspIleThrTrpValGl 375

 1308 CGCGGGCGTTCATATCAGTGATGGCAGTACCGTTACTTGGAAAGTAAACG 1357
      |:::|||:::   :::::::::|||:::   |||:::|||||||||:::::
  375 yGlyGlyIleAspValAlaGluGlyLysGluValValTrpLysValHisA 392

 1358 GCGTGGCAAACGACCGCCTGTCCAAAATCGGCAAAGGCACGCTGCTGGTT 1407
      ::      :::|||:::|||:::|||||||||||||||||||||:::|||
  392 snProGluLysAspHisLeuAlaLysIleGlyLysGlyThrLeuIleVal 408

 1408 CAAGCCAAAGGGGAAAACCAAGGCTCGGTCAGCGTGGGCGACGGTAAAGT 1457
      ::::::|||||:::|||:::||||||:::::|||||||||||   ||
  409 GluGlyLysGlyAsnAsnLysGlySerLeuLysValGlyAspGlyThrVa 425

 1458 CATCTTAGATCAGCAGGCGGACGATCAAGGCAAAAAACAAGCCTTTAGTG 1507
      |:::|||   ||||||::::::   :::|||   ::::::||||||::::
  425 lValLeuLysGlnGlnThrAsnGlySerGly...GluHisAlaPheAlaS 441

 1508 AAATCGGCTTGGTCAGCGGCAGGGGGACGGTGCAACTGAATGCCGATAAT 1557
      :::::|||:::|||||||||||||:::||||||   ||||||   |||:::
  441 erValGlyIleValSerGlyArgSerThrValValLeuAsnAspAspLys 457

 1558 CAGTTCAACCCCGACAAACTCTATTTCGGCTTTCGCGGCGGA 1599
      |||   :::|||::::::::::|||||||||||||||||||
  458 GlnValAspProAsnSerIleTyrPheGlyPheArgGlyGly 471
```

seq_name: pir2:T00317

seq_documentation_block:
probable serine proteinase espP, extracellular - Escherichia coli plasmid pO157
N;Alternate names: putative exoprotein-precursor
C;Species: Escherichia coli
C;Date: 01-Feb-1999 #sequence_revision 01-Feb-1999 #text_change 08-Dec-2000
C;Accession: T00317; T42120
R;Makino, K.; Ishii, K.; Yasunaga, T.; Hattori, M.; Yokoyama, K.; Yatsudo, H.C.; Kubota,
S.; Shinagawa, H.
DNA Res. 5, 1-9, 1998
A;Title: Complete nucleotide sequences of 93-kb and 3.3-kb plasmids of an enterohemorrha
A;Reference number: Z14127; MUID:98290540
A;Accession: T00317
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-1300 <MAK>
A;Cross-references: EMBL:AB011549; NID:g4589740; PIDN:BAA31836.1; PID:g3337077
A;Experimental source: strain EHEC O157:H7, substrain RIMD 0509952
R;Burland, V.; Shao, Y.; Perna, N.T.; Plunkett, G.; Sofia, H.J.; Blattner, F.R.
Nucleic Acids Res. 26, 4196-4204, 1998
A;Title: The complete DNA sequence and analysis of the large virulence plasmid of Escher
A;Reference number: Z22068; MUID:98391744
A;Accession: T42120
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-1300 <BUR>
A;Cross-references: EMBL:AF074613; PIDN:AAC70088.1
A;Experimental source: strain EDL933; serotype O157:H7
C;Genetics:
A;Gene: espP
A;Genome: plasmid pO157
C;Superfamily: IgA-specific metalloendopeptidase
C;Keywords: serine proteinase

alignment_scores:
Quality: 629.50 Length: 1505
Ratio: 0.844 Gaps: 64
Percent Similarity: 49.568 Percent Identity: 21.794

alignment_block:
US-09-303-518D-653 x T00317 ..

Align seg 1/1 to: T00317 from: 1 to: 1300

```
   82 TTAGCCATATGCCTGTCGTTCGGCATTCTGCCCCAAGCCCGGGCGGGACA 131
      ||||||:::|||:::   :::|||:::|||   ::::::   :::
   40 LeuAlaLeuCysPhe...LeuGlyLeuLeuGlnSerSerTyrSerPheAl 55

  132 CACTTATTTCGGCATCAACTACCAATAC.........TATCGCGACTTTG 172
       :::   :::   |||:::      |||         |||   |||||||
   55 aSerGlnMetAspIleSerAsnPheTyrIleArgAspTyrMetAspPheA 72

  173 CCGAAAATAAAGGCAAGTTTGCAGTCGGGGCGAAAGATATTGAGGTTTAC 222
      ||:::|||||||||   |||   :::||||||   :::|||||||:::
   72 laGlnAsnLysGlyIlePheGlnAlaGlyAlaThrAsnIleGluIleVal 88

  223 AACAAAAAAGGGGAGTTGGTCGGCAAATCGATGACGAAAGCCCCGATGAT 272
      :::|||   |||:::   :::      ::::::   :::::::|||:::
   89 LysLysAspGlySerThrLeu......LysLeuProGluValProPhePr 103

  273 TGATTTTTCTGTGGTATCGCGTAACGGCGTGGCGGCATTGGCGGGCGATC 322
       |||||||||   |||::::::::::|||   ::::::      |||
  103 oAspPheSerProValAlaAsnLysGlySerThrThrSerIleGlyGlyA 120

  323 AATATATTGTGAGCGTGGCACATAACGGCGGCTATAACAATGTTGATTTT 372
        |||   :::::::::::::||||||
  120 laTyrSerIleThrAlaThrHisAsn..................... 128

  373 GGTGCGGAGGGAAGCAATCCCGATCAGCACCGCTTTTCTTACCAAATTGT 422
                      :::   ::::::|||      :::   |||
  129 ..............ThrLysAsnHisHisSerValAlaThrGlnAsnTr 140

  423 GAAAAGAAATAATTATAAAGCA......GGGACTAACGGCCATCCTTATG 466
          :::::::::||||||            :::::::::||||||
  140 pGlyAsnSerThrTyrLysGlnThrAspTrpAsnThrSerHisPro.... 155

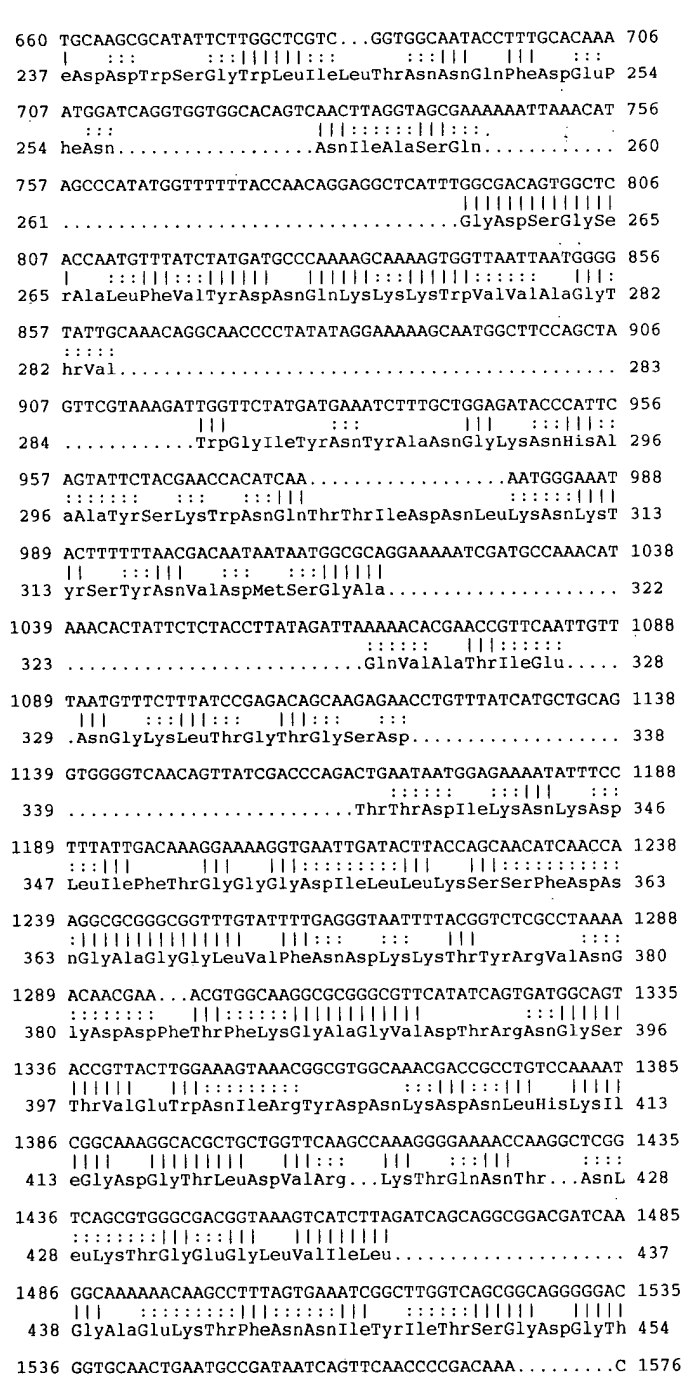
  467 GCGGCGATTATCATATGCCGCGTTTGCACAAATTTGTCACAGATGCAGAA 516
          |||:::   :::   ||||||   |||||||||:::::::::::
  156 .....AspPheAlaValSerArgLeuAspLysPheValValGluThrArg 170

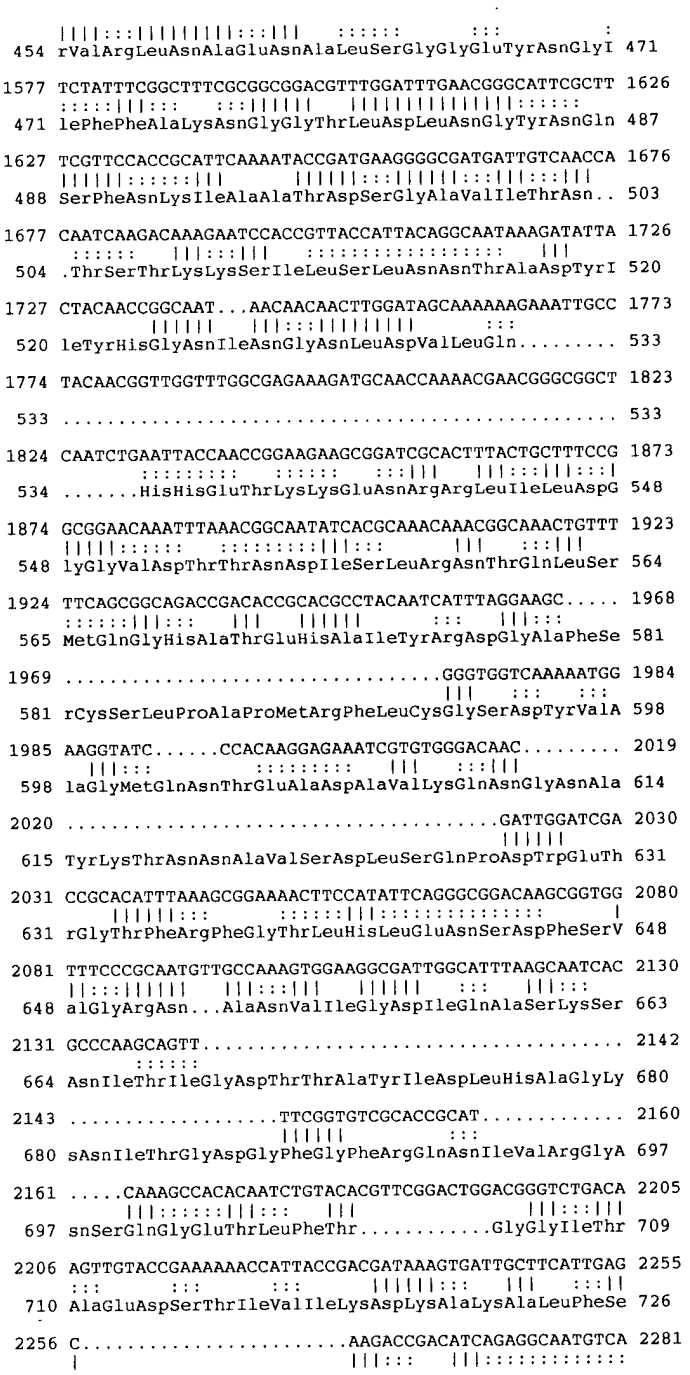
  517 CCTGTTGAGATGACCAGTTATATGGATGGGTGGAAATACGCTGATTTAAA 566
        :::      :::   :::::      |||         |||::
  171 GlyAlaThrGluGlyAlaAspIleSerLeuSerLysGlnGlnAlaLeuGl 187

  567 TAAATACCCTGATCGT.........................GTTCGAA 589
      ::::|||      :::                          |||
  187 uArgTyrGlyValAsnTyrLysGlyGluLysLysLeuIleAlaPheArgA 204

  590 TCGGAGCAGGCAGACAATATTGGCGGTCTGATGAAGACGAACCCAATAAC 639
        |||:::|||           :::::::::
  204 laGlySerGlyValValSerValLysLysAsnGlyArgIleThrProPhe 220

  640 CGCGAAAGTTCATAT.............................CATAT 659
      :::|||   ||||||                             |||||
  221 AsnGluValSerTyrLysProGluMetLeuAsnGlySerPheValHisIl 237
```


660 TGCAAGCGCATATTCTTGGCTCGTC...GGTGGCAATACCTTTGCACAAA 706
237 eAspAspTrpSerGlyTrpLeuIleLeuThrAsnAsnGlnPheAspGluP 254
707 ATGGATCAGGTGGTGGCACAGTCAACTTAGGTAGCGAAAAAATTAAACAT 756
254 heAsn.....AsnIleAlaSerGln..... 260
757 AGCCCATATGGTTTTTTACCAACAGGAGGCTCATTTGGCGACAGTGGCTC 806
261 .....GlyAspSerGlySe 265
807 ACCAATGTTTATCTATGATGCCCAAAAGCAAAAGTGGTTAATTAATGGGG 856
265 rAlaLeuPheValTyrAspAsnGlnLysLysLysTrpValValAlaGlyT 282
857 TATTGCAAACAGGCAACCCCTATATAGGAAAAAGCAATGGCTTCCAGCTA 906
282 hrVal..... 283
907 GTTCGTAAAGATTGGTTCTATGATGAAATCTTTGCTGGAGATACCCATTC 956
284 .....TrpGlyIleTyrAsnTyrAlaAsnGlyLysAsnHisAl 296
957 AGTATTCTACGAACCACATCAA.....AATGGGAAAT 988
296 aAlaTyrSerLysTrpAsnGlnThrThrIleAspAsnLeuLysAsnLysT 313
989 ACTTTTTTAACGACAATAATAATGGCGCAGGAAAAATCGATGCCAAACAT 1038
313 yrSerTyrAsnValAspMetSerGlyAla..... 322
1039 AAACACTATTCTCTACCTTATAGATTAAAAACACGAACCGTTCAATTGTT 1088
323 .....GlnValAlaThrIleGlu..... 328
1089 TAATGTTTCTTTATCCGAGACAGCAAGAGAACCTGTTTATCATGCTGCAG 1138
329 .AsnGlyLysLeuThrGlyThrGlySerAsp..... 338
1139 GTGGGGTCAACAGTTATCGACCCAGACTGAATAATGGAGAAAATATTTCC 1188
339 .....ThrThrAspIleLysAsnLysAsp 346
1189 TTTATTGACAAAGGAAAAGGTGAATTGATACTTACCAGCAACATCAACCA 1238
347 LeuIlePheThrGlyGlyGlyAspIleLeuLeuLysSerSerPheAspAs 363
1239 AGGCGCGGGCGGTTTGTATTTTGAGGGTAATTTTACGGTCTCGCCTAAAA 1288
363 nGlyAlaGlyGlyLeuValPheAsnAspLysLysThrTyrArgValAsnG 380
1289 ACAACGAA...ACGTGGCAAGGCGCGGGCGTTCATATCAGTGATGGCAGT 1335
380 lyAspAspPheThrPheLysGlyAlaGlyValAspThrArgAsnGlySer 396
1336 ACCGTTACTTGGAAAGTAAACGGCGTGGCAAACGACCGCCTGTCCAAAAT 1385
397 ThrValGluTrpAsnIleArgTyrAspAsnLysAspAsnLeuHisLysIl 413
1386 CGGCAAAGGCACGCTGCTGGTTCAAGCCAAAGGGGAAAACCAAGGCTCGG 1435
413 eGlyAspGlyThrLeuAspValArg...LysThrGlnAsnThr...AsnL 428
1436 TCAGCGTGGGCGACGGTAAAGTCATCTTAGATCAGCAGGCGGACGATCAA 1485
428 euLysThrGlyGluGlyLeuValIleLeu..... 437
1486 GGCAAAAAACAAGCCTTTAGTGAAATCGGCTTGGTCAGCGGCAGGGGGAC 1535
438 GlyAlaGluLysThrPheAsnAsnIleTyrIleThrSerGlyAspGlyTh 454
1536 GGTGCAACTGAATGCCGATAATCAGTTCAACCCCGACAAA.....C 1576


454 rValArgLeuAsnAlaGluAsnAlaLeuSerGlyGlyGluTyrAsnGlyI 471
1577 TCTATTTCGGCTTTCGCGGCGGACGTTTGGATTTGAACGGGCATTCGCTT 1626
471 lePhePheAlaLysAsnGlyGlyThrLeuAspLeuAsnGlyTyrAsnGln 487
1627 TCGTTCCACCGCATTCAAAATACCGATGAAGGGGCGATGATTGTCAACCA 1676
488 SerPheAsnLysIleAlaAlaThrAspSerGlyAlaValIleThrAsn.. 503
1677 CAATCAAGACAAAGAATCCACCGTTACCATTACAGGCAATAAAGATATTA 1726
504 .ThrSerThrLysLysSerIleLeuSerLeuAsnAsnThrAlaAspTyrI 520
1727 CTACAACCGGCAAT...AACAACAACTTGGATAGCAAAAAAGAAATTGCC 1773
520 leTyrHisGlyAsnIleAsnGlyAsnLeuAspValLeuGln..... 533
1774 TACAACGGTTGGTTTGGCGAGAAAGATGCAACCAAAACGAACGGGCGGCT 1823
533 533
1824 CAATCTGAATTACCAACCGGAAGAAGCGGATCGCACTTTACTGCTTTCCG 1873
534 .....HisHisGluThrLysLysGluAsnArgArgLeuIleLeuAspG 548
1874 GCGGAACAAATTTAAACGGCAATATCACGCAAACAAACGGCAAACTGTTT 1923
548 lyGlyValAspThrThrAsnAspIleSerLeuArgAsnThrGlnLeuSer 564
1924 TTCAGCGGCAGACCGACACCGCACGCCTACAATCATTTAGGAAGC..... 1968
565 MetGlnGlyHisAlaThrGluHisAlaIleTyrArgAspGlyAlaPheSe 581
1969 .....GGGTGGTCAAAAATGG 1984
581 rCysSerLeuProAlaProMetArgPheLeuCysGlySerAspTyrValA 598
1985 AAGGTATC.....CCACAAGGAGAAATCGTGTGGGACAAC..... 2019
598 laGlyMetGlnAsnThrGluAlaAspAlaValLysGlnAsnGlyAsnAla 614
2020 .....GATTGGATCGA 2030
615 TyrLysThrAsnAsnAlaValSerAspLeuSerGlnProAspTrpGluTh 631
2031 CCGCACATTTAAAGCGGAAAACTTCCATATTCAGGGCGGACAAGCGGTGG 2080
631 rGlyThrPheArgPheGlyThrLeuHisLeuGluAsnSerAspPheSerV 648
2081 TTTCCCGCAATGTTGCCAAAGTGGAAGGCGATTGGCATTTAAGCAATCAC 2130
648 alGlyArgAsn...AlaAsnValIleGlyAspIleGlnAlaSerLysSer 663
2131 GCCCAAGCAGTT..... 2142
664 AsnIleThrIleGlyAspThrThrAlaTyrIleAspLeuHisAlaGlyLy 680
2143 .....TTCGGTGTCGCACCGCAT..... 2160
680 sAsnIleThrGlyAspGlyPheGlyPheArgGlnAsnIleValArgGlyA 697
2161 .....CAAAGCCACACAATCTGTACACGTTCGGACTGGACGGGTCTGACA 2205
697 snSerGlnGlyGluThrLeuPheThr.....GlyGlyIleThr 709
2206 AGTTGTACCGAAAAAACCATTACCGACGATAAAGTGATTGCTTCATTGAG 2255
710 AlaGluAspSerThrIleValIleLysAspLysAlaLysAlaLeuPheSe 726
2256 C.....AAGACCGACATCAGAGGCAATGTCA 2281

```
 726 rAsnTyrValTyrLeuLeuAsnThrLysAlaThrIleGluAsnGlyAlaA 743

2282 GCCTTGCCGATCACGCTCATTTAAATCTCACAGGACTTGCCACACTCAAC 2331
743 spValThrThrGlnSerGlyMetPheSerThrSerAspIleSerIleSer 759

2332 GGCAATCTTAGTGCAGGCGGA..........................  2352
760 GlyAsnLeuSerMetThrGlyAsnProAspLysAspAsnLysPheGluPr 776

2353 ..............GACACGCACTATACGGTTACGCGCAACGCCACCC 2386
776 oSerIleTyrLeuAsnAspAlaSerTyrLeuLeuThrAspAspSerAlaA 793

2387 AAAACGGCAACCTCAGCCTCGTGGGCAATGCCCAAGCAACATTTAATCAA 2436
793 rg..............LeuValAlaLys.............AsnLys 799

2437 GCCACATTAAACGGCAACACATCGGCTTCGGACAATGCTTCATTTAATCT 2486
800 AlaSerValValGlyAspIleHisSerThrLysSerAlaSerIleMetPh 816

2487 AAGCAACAACGCCGTACAAAACGGCAGTCTGACGCTTTCCGACAACGCTA 2536
816 eGlyHisAsp......GluSerAspLeuSerGlnLeuSerAspArgThrS 831

2537 AGGCAAACGTAAGCCATTCCGCACTCAACGGCAATGTCTCCCTAGCCGAT 2586
831 erLysGlyLeuAlaLeuGlyLeuLeu..................... 839

2587 AAGGCAGTATTCCATTTTGAAAACAGCCGCTTTACCGGAAAAATCAGCGG 2636
840 .....................................................Gl 840

2637 CGGCAAGGATACGGCATTACACTTAAAAGACAGCGAATGGACGCTGCCGT 2686
840 yGlyPheAspValSerTyrArg.......................... 847

2687 CGGGCACGGAATTAGGCAATTTAAACCTTGACAACGCCACCATTACACTC 2736
848 ..............GlySerValAsnAlaProSerAlaSerAlaThrMet 859

2737 AATTCCGCCTATCGACACGATGCGGCAGGCGCGCAAACCGGCAGTGCGGC 2786
860 AsnAsnThrTrp..............TrpGlnLeuThrGlyAspSerAl 871

2787 AGATGCGCCGCGCCGCCGTTCGCGCCGTTCCCTATTATCCGTTACGCCGC 2836
871 a...LeuLysThrLeuLysSerThrAsnSerMetValTyrPheThr...A 886

2837 CAACTTCGGCAGAATCCCGTTTCAACACGCTGACGGTAAACGGCAAATTG 2886
886 spSerAlaAsnAsnLysLysPheHisThrLeuThrValAspGluLeuAla 902

2887 AACGGTCAGGGAACATTCCGCTTTATGTCGGAACTCTTCGGCTACCGCAG 2936
903 ThrSerAsnSerAlaTyrAlaMetArgThrAsnLeu......SerGluSe 917

2937 CGGCAAATTGAAGCTGGCGGAAAGTTCCGAAGGCACTTACACCTTGGCTG 2986
917 rAspLysLeuGluValLysLysHisLeuSerGlyGluAsnAsnIleLeuL 934

2987 TCAACAATACCGGCAACGAACCCGTAAGTCTCGAGCAATTGACGGTA... 3033
934 euValAspPheLeuGlnLysProThrProGluLysGlnLeuAsnIleGlu 950

3034 ...GTGGAAGGAAAAGACAACACACCGCTGTCCGAAAATCTTAATTTCAC 3080
951 LeuValSerAlaProLysAspThr........................ 958

3081 CCTGCAAAACGAACACGTCGATGCCGGCGCATGGCGTTATCAGCTTATCC 3130
959 .......AsnGluAsnVal.................PheLysAlaSerL 967

3131 GCAAAGACGGCGAGTTCCGCCTGCATAATCCGGTC...AAAGAACAAGAG 3177
967 ysGlnThrIleGlyPheSerAspValThrProValIleThrThrArgGlu 983

3178 CTTTCCGACAAACTC.....................GGCAAGGCGGGAGA 3206
984 ThrAspAspLysIleThrTrpSerLeuThrGlyTyrAsnThrValAlaAs 1000

3207 AACAGAGGCCGCCTTGACGGCAAAACAGGCACAACTTGCCGCCAAACAAC 3256
1000 nLysGluAlaThrArgAsnAla...AlaAlaLeuPheSerValAspTyrL 1016

3257 AGGCGGAAAAAGACAACGCGCAAAGCCTTGACGCGCTGATTGCGGCCGGG 3306
1016 ysAlaPheLeuAsnGluValAsnAsnLeuAsnLysArgMetGlyAspLeu 1032

3307 CGCAATGCCACCGAAAAGGCAGAAAGTGTTGCCGAACCGGCCCGGCAGGC 3356
1033 ArgAspIleAsnGlyGluAlaGlyAlaTrpAlaArgIleMetSerGlyTh 1049

3357 AGGCGGGGAAAATGCCGGCATTATGCAGGCGGAGGAAGAGAAAAAACGGG 3406
1049 rGlySerAlaSerGlyGlyPhe.........SerAspAsnTyrThrHisV 1063

3407 TGCAGGCGGATAAAGACACCGCCTTGGCGAAACAGCGCGAAGCGGAAACC 3456
1063 alGlnValGlyValAsp............LysLysHisGluLeuAspGly 1075

3457 CGGCCGGCTACCACCGCCTTCCCCCGCGCCCGCCGCGCCCCGCCGGGATTT 3506
1076 LeuAspLeuPheThrGlyPhe........................... 1082

3507 GCCGCAACCGCAGCCCCAACCGCAACCCCAACCGCAGCGCGACCTGATCA 3556
1083 .......ThrValThrHisThrAspSerSerAlaSerAlaAspValPheS 1097

3557 GCCGTTATGCCAATAGC...GGTTTGAGTGAATTTTCCGCCACGCTCAAC 3603
1097 erGlyLysThrLysSerValGlyAlaGlyLeuTyrAlaSerAlaMetPhe 1113

3604 AGCGTTTTCGCCGTACAGGACGAATTGGACCGCGTGTTTGCCGAAGACCG 3653
1114 AspSerGlyAlaTyrIleAspLeuIleGlyLysTyrValHisHisAspAs 1130

3654 CCGCAACGCCGTTTGGACAAGCGGCATCCGGGACACCAAACACTACCGTT 3703
1130 nGluTyrThrAlaThrPheAlaGlyLeu...GlyThrArgAspTyrSerT 1146

3704 CGCAAGATTTCCGCGCC..............TACCGCCAACAA...ACC 3735
1146 hrHisSerTrpTyrAlaGlyAlaGluAlaGlyTyrArgTyrHisValThr 1162

3736 GACCTGCGCCAAATCGGTATGCAGAAAAACCTCGGCAGCGGGCGCGTCGG 3785
1163 GluAspAlaTrpIleGluProGlnAlaGluLeuValTyrGlySerVal.. 1178

3786 CATCCTGTTTTCGCACAACCGGACCGGAAACACCTTC.........GACG 3826
1179 .....................SerGlyLysGlnPheAlaTrpLysAspG 1188

3827 ACGGCATCGGCAACTCGGCACGG...CTTGCCCACGGTGCCGTTTTCGGG 3873
1188 lnGlyMetHisLeuSerMetLysAspLysAspTyrAsnProLeuIleGly 1204

3874 CAATACGGCATCGGCAGGTTCGACATCGGCATCAGC.............. 3909
1205 ArgThrGlyVal.........AspValGlyLysSerPheSerGlyLysAs 1218

3910 ..................GCGGGCGCGGGTTTTAGTAGCGGCAGCCTTT 3940
1218 pTrpLysValThrAlaArgAlaGlyLeuGlyTyrGlnPheAspLeuLeuA 1235
```

```
   3941 CAGACGGC.................................ATCAGA 3954
        :::::|||                                 |||:::
   1235 laAsnGlyGluThrValLeuArgAspAlaSerGlyGluLysArgIleLys 1251

   3955 GGC...AAAATCCGCCGCCGCGTGCTGCATTACGGCATTCAGGCAAGATA 4001
        |||   |||      |||   ::::::      |||::::::|||:::
   1252 GlyGluLysAspSerArgMetLeuMetSerValGlyLeuAsnAlaGluIl 1268

   4002 CCGC..............................GCAGGTTTCGGCG 4018
         |||                             ::::::||||||
   1268 eArgAspAsnValArgPheGlyLeuGluPheGluLysSerAlaPheGlyL 1285

   4019 GATTCGGCATCGAACCGCACATCGGCGCAACGCGCTATTTCGTCCAAAAA 4068
          ::::::::::::      :::                    :::
   1285 ysTyrAsnValAspAsnAlaVal.....................Asn 1293

   4069 GCGGATTACCGATAC 4083
        |||::::::|||||||
   1294 AlaAsnPheArgTyr 1298

seq_name: pir2:B41500

seq_documentation_block:
IgA-specific metalloendopeptidase (EC 3.4.24.13) type 2 - Haemophilus influenzae (fragme
N;Alternate names: immunoglobulin A1 proteinase type 2
C;Species: Haemophilus influenzae
C;Date: 30-Jun-1992 #sequence_revision 30-Jun-1992 #text_change 08-Dec-2000
C;Accession: B41500
R;Grundy, F.J.; Plaut, A.G.; Wright, A.
Infect. Immun. 58, 320-331, 1990
A;Title: Localization of the cleavage site specificity determinant of Haemophilus influe
A;Reference number: A41500; MUID:90129281
A;Accession: B41500
A;Status: not compared with conceptual translation
A;Molecule type: DNA
A;Residues: 1-384 <GRU>
C;Comment: This IgA1 proteinase is classified as type 2 because it cleaves at a proline-
C;Superfamily: IgA-specific metalloendopeptidase
C;Keywords: hydrolase; metalloproteinase


alignment_scores:
            Quality:  407.00              Length:    429
              Ratio:  1.668                Gaps:     21
 Percent Similarity:  56.876   Percent Identity: 29.138

alignment_block:
 US-09-303-518D-653 x B41500   ..

 Align seg 1/1  to: B41500  from: 1  to: 384

      64 CGCTTCTCGCCCGCTTACTTAGCCATATGCCTGTCGTTCGGCATTCTGCC 113
         :::|||:::      :::::::|||:::   ::::::::::::::   ||
       5 LysPheLysLeuAsnPheIleAlaLeuThrValAlaTyrAlaLeuThrPr 21

     114 CCAAGCCCGGGCGGGACACACTTATTTCGGCATCAACTACCAATACTATC 163
         |   ::::::|||:::   :::         ::::::||||||   :::|
      21 oTyrThrGluAlaAlaLeuValArgAspAspValAspTyrGlnIlePheA 38

     164 GCGACTTTGCCGAAAATAAAGGCAAGTTTGCAGTCGGGGCGAAAGATATT 213
         ||||||||||||||||||||||:::|||:::|||||||||   ::::::
      38 rgAspPheAlaGluAsnLysGlyArgPheSerValGlyAlaThrAsnVal 54

     214 GAGGTTTACAACAAAAAAGGGGAGTTGGTCGGCAAATCGATGACGAAAGC 263
         ||||||   :::|||:::::::::   :::|||:::   :::   :::::
      55 GluValArgAspLysAsnAsnHisSerLeuGlyAsnValLeuProAsnGl 71

     264 C...CCGATGATTGATTTTTCTGTGGTATCGCGTAAC...GGCGTGGCGG 307
         :   |||||||||||||||||||||||:::   |||      :::::::
      71 yIleProMetIleAspPheSerValValAspValAsnLysArgIleGlyT 88

     308 CATTGGCGGGCGATCAATATATTGTGAGCGTGGCACATAACGGCGGCTAT 357
         ::|||:::      ||||||||||||||||||   |||         |||

      88 hrLeuValAspProGlnTyrIleValSerValLysHisAlaHisGlnTyr 104

     358 AACAATGTTGATTTT.................GGTGCGGAGGGAAGCAA 389
           |||   ||||||                   |||   :::      ::
     105 MetAsn...AspPheTyrPheGlyHisTyrAsnGlyHisArgAspValSe 120

     390 TCCCGATCAGCACCGCTTTTCTTACCAAATTGTGAAAAGAAATAATTATA 439
         :   |||:::::::::      |||:::::|||   :::||||||   :
     120 rAspAspGluAsnLys......TyrSerValValThrGlnAsnAsnValA 135

     440 AAGCAGGGACTAACGGCCAT......CCTTATGGCGGCGATTATCATATG 483
         ::   :::   |||   |||                  ||||||:::|||
     135 snProAsnGluAsnTrpHisValAspLysArgLeuAspAspTyrAsnMet 151

     484 CCGCGTTTGCACAAATTTGTCACAGATGCAGAACCTGTTGAGATGACCAG 533
         |||||||||:::|||||||||||||||:::::   |||:::      |||
     152 ProArgLeuAsnLysPheValThrGluValAlaProThrThrProThrLe 168

     534 TTATATGGATGGGTGGAAA...TACGCTGATTTAAATAAATACCCTGATC 580
               |||      :::   |||   |||   :::||||||   :::
     168 uAlaGlyAspAspLeuGluThrTyrLysAspLysGluLysTyrLeuSerP 185

     581 GTGTTCGAATCGGAGCAGGCAGACAA......TATTGGCGGTCTGATGAA 624
         ||||||:::|||||||||||||||      |||   ::::::::::::
     185 heValArgValGlyAlaGlyArgGlnLeuValTyrGluLysGlySerHis 201

     625 ......GACGAACCCAATAACCGCGAAAGTTCATATCATATTGCAAGCGC 668
               ::::::   ::::::  |||:::          :::::::::||
     202 HisValGluAspLysGluHisGlyGluAspLeuLysAspLeuSerAlaAl 218

     669 ATATTCTTGGCTCGTCGGTGGCAATACCTTTGCA.............C 703
         ||||   :::   :::|||||||:::   :::
     218 aTyrArgTyrAlaIleGlyGlyThrProTyrLysGlyIleAsnIleAspP 235

     704 AAAATGGATCAGGTGGTGGCACAGTCAACTTAGGTAGCGAAAAAATTAAA 753
           :::   |||      |||   :::::::::|||::::::::::
     235 roSerGlnSerLysLysGlyLeuIleGlyPheGlyAspSerArgGluAsp 251

     754 CAT.....................................AGCCCATA 764
         |||                                     :::   ||
     252 HisValIleAsnSerLysThrLeuLeuSerGlnAspProLeuThrAsnTy 268

     765 TGGTTTTTTACCAACAGGAGGCTCATTTGGCGACAGTGGCTCACCAATGT 814
         ||||   |||                 ||||||||||||||||||:::|
     268 rGlyValLeu.................GlyAspSerGlySerProLeuP 279

     815 TTATCTATGATGCCCAAAAGCAAAAGTGGTTAATTAATGGGGTATTGCAA 864
         ||   :::|||   |||::::::|||||||::::::   |||
     279 heAlaPheAspLysGlnGlnAsnLysTrpPhePheIleGly..... 292

     865 ACAGGCAACCCCTAT..............ATAGGAAAAAGCAATGGCTT 899
                |||||                 ||||||::::::
     293 .........ProTyrThrTyrTrpAlaGlyTyrGlyLysLysSer..... 304

     900 CCAGCTAGTTCGTAAAGATTGG...TTCTATGATGAAATCTTTGCTGGAG 946
                       :::::::|||   :::|||   :::   |||:::   |
     305 ..........ThrGlnGluTrpAsnIleTyrLysSerGlnPheThrLysA 318

     947 ATACCCATTCAGTATTCTACGAACCACATCAAAATGGGAAATACTTTTTT 996
         ||:::
     318 spValLeu..................................... 320

     997 AACGACAATAATAATGGCGCAGGAAAAATCGATGCCAAACATAAACACTA 1046
               |||::::::::::||||||   :::   :::   |||   :::||
     321 ......AsnLysAspSerAlaGlyLeuLeuLysGly...HisThrGlnTy 334

    1047 TTCTCTACCTTATAGATTAAAAACACGAACCGTT............CAAT 1084
         |:::         :::   :::|||      :::            :::|
     334 rAsnTrpThrSerAsnGlyAsnThrSerMetIleSerAsnGlySerGluL 351

    1085 TGTTTAATGTTTCTTTATCCGAGACAGCAAGAGAACCTGTTTATCATGCT 1134
         ||::::::|||:::|||   ::::::::::::
     351 euLeuGlyValAsnLeuPheAspAsnSerLys.................. 361
```

```
 1135 GCAGGTGGGGTCAACAGTTATCGACCCAGACTGAATAATGGAGAAAATAT 1184
                :::::   |||   :::   |||   |||::::::::
  362 ...........HisThrAsnArgGluLysAlaAsnTyrGlyLysSerVa 374

 1185 TTCCTTTATTGACAAAGGAAAAGGTGAATTGATACTT 1221
      ::::|||      :::|||:::|||   |||   |||
  374 lThrPhe......GlnGlyAsnGlyThrLeuThrLeu 384

seq_name: pir2:A65044

seq_documentation_block:
hypothetical protein b2647 - Escherichia coli (strain K-12)
C;Species: Escherichia coli
C;Date: 12-Sep-1997 #sequence_revision 17-Sep-1997 #text_change 08-Oct-1999
C;Accession: A65044
R;Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Co
.A.; Rose, D.J.; Mau, B.; Shao, Y.
Science 277, 1453-1462, 1997
A;Title: The complete genome sequence of Escherichia coli K-12.
A;Reference number: A64720; MUID:97426617
A;Accession: A65044
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-1569 <BLAT>
A;Cross-references: GB:AE000350; GB:U00096; NID:g2367147; PIDN:AAC75695.1; PID:g1789001;
A;Experimental source: strain K-12, substrain MG1655


alignment_scores:
           Quality:  359.00            Length:   1686
             Ratio:  0.481               Gaps:     80
 Percent Similarity:  44.306   Percent Identity: 19.395

alignment_block:
 US-09-303-518D-653 x A65044   ..

 Align seg 1/1  to: A65044  from: 1  to: 1569

  445 GGGACTAACGGCCATCCTTATGGCGGCGATTATCATATGCCGCGTTTGCA 494
      ||||||:::::::::   :::      ::::::   :::   :::::::||
   97 GlyThrThrAsnAsnThrHisIleIleAsnHisGlyGlnGlnAsnValHi 113

  495 CAAATTTGTCACAGATGCAGAACCTGTTGAGATGACCAGTTATATGGATG 544
      |      |||:::::::::::::   :::|||      :::|||:::|||
  113 sGlyGlyValSerAsnGlySerLeuIleGluSerGlyGlyTyrGlnAspI 130

  545 GGTGGAAATACGCTGATTTAAATAAATACCCTGATCGTGTTCGAATCGGA 594
          ::::::   ::::::            :::::::::   |||:::
  130 leGlySerHisAsnAsnPheValGlyGlnAlaAsnAsnThrThrIleAsn 146

  595 GCAGGCAGACAATATTGGCGGTCTGATGAAGACGAACCCAATAACCGCGA 644
      :::|||||||||
  147 GlyGlyArgGln..................................... 150

  645 AAGTTCATATCATATTGCA......AGCGCATATTCTTGGCTCGTCGGTG 688
          |||   |||   :::      |||:::   :::   :::   :::|
  151 ....SerIleHisAspGlyGlyIleSerThrGlyThrThrIleGluSerG 166

  689 GCAATACCTTTGCACAAAATGGA......TCAGGTGGTGGCACAGTCAAC 732
      |||||      ::   :::|||      |||:::|||   |||::::::
  166 lyAsnGlnAspValTyrLysGlyGlyIleSerAsnGlyThrThrIleLys 182

  733 TTAGGTAGCGAAAAAATTAAA...CATAGCCCATATGGTTTTTTACCAAC 779
          |||:::::::::::::      |||      |||:::|||
  183 GlyGlyAlaSerArgValGluGlyGlySerAlaAsnGlyIleLeuIleAs 199

  780 AGGAGGCTCA.....................................T 790
       |||||||||                                     :
  199 pGlyGlySerGlnIleValLysValGlnGlyHisAlaAspGlyThrThrI 216

  791 TTGGCGACAGTGGCTCACCAATGTTTATCTATGATGCCCAAAAGCAAAAG 840
      :::::   |||||||||         :::      :::         :::

  216 leAsnLysSerGlySerGlnAspValValGlnGlySerLeuAlaThrAsn 232

  841 TGGTTAATTAATGGGGTATTGCAAACAGGCAACCCCTATATAGGA..... 885
          |||||||||             |||:::   |||:::
  233 ThrThrIleAsnGly............GlyArgGlnTyrValGluGlnSe 245

  886 .....................AAAAGCAATGGCTTCCAGCTAGTTCGTA 913
                              |||:::::::|||   |||   |||  :
  245 rThrValGluThrThrThrIleLysAsnGlyGlyGluGlnArgValTyrG 262

  914 AAGATTGGTTCTATGATGAAATCTTTGCTGGAGATACCCATTCAGTATTC 963
      :::::         |||      :::   |||   |||:::|||:::
  262 luSerArgAlaLeuAspThrThrIleGluGlyGlyThrGlnSerLeu... 277

  964 TACGAACCACATCAAAATGGGAAATACTTTTTTAACGACAATAATAATGG 1013

  277 ..................................................... 277

 1014 CGCAGGAAAAATCGATGCCAAACATAAACACTATTCTCTACCTTATAGAT 1063
          :::|||      ||||||:::   |||             |||
  278 .AsnSerLysSerThrAlaLysAsnThrHisIle.........TyrSerG 291

 1064 TAAAAACACGAACCGTTCAATTGTTTAATGTTTCTTTATCCGAGACAGCA 1113
          |||:::   |||:::      |||:::|||   |||::::::
  291 lyGlyThrGlnIleValAsp......AsnThrSerThrSerAspValIle 305

 1114 AGAGAACCTGTTTATCATGCTGCAGGTGGGGTCAACAGTTATCGACCCAG 1163
      :::      ||||||      :::||||||||||   :::   |||
  306 Glu......ValTyr......SerGlyGlyValLeuAspValArg..... 316

 1164 ACTGAATAATGGAGAAAATATTTCCTTTATTGACAAAGGAAAAGGTGAAT 1213
               |||   :::   :::   :::   :::      |||   :
  317 ..........GlyGlyThrAlaThrAsnValThrGlnHisAspGlyAlaI 330

 1214 TGATACTTACCAGCAACATCAACCAAGGCGCGGGCGGTTTGTATTTTGAG 1263
      :::::   |||::::::   :::      :::::|||         |||
  330 leLeuLysThrAsnThrAsnGlyThrThrValSerGlyThrAsnSerGlu 346

 1264 GGTAATTTTACGGTCTCGCCTAAAAACAACGAAACGTGGCAAGGCGCGGG 1313
      |||   |||::::::         :::|||::::::      :::   ::
  347 GlyAlaPheSerIle.........HisAsnHisVal......AlaAspAs 358

 1314 CGTTCATATCAGTGATGGCAGTACCGTTACTTGGAAAGTAAACGGCGTGG 1363
      :|||   :::::::::|||:::   :::      ::::::   |||   |
  358 nValLeuLeuGluAsnGlyGlyHisLeuAspIleAsnAlaTyrGlySerA 375

 1364 CAAACGACCGCCTGTCCAAAATCGGCAAAGGCACGCTG......CTGGTT 1407
      |||||      :::   |||      |||||||||:::      |||:::
  375 laAsnLysThrIleIleLys...AspLysGlyThrMetSerValLeuThr 390

 1408 CAAGCCAAAGGG..............GAAAACCAAGGCTCGGTCAGCGT 1442
      :::||||||:::              :::|||   |||   ::::::||
  391 AsnAlaLysAlaAspAlaThrArgIleAspAsnGlyGlyValMetAspVa 407

 1443 GGGCGACGGT......AAAGTCATCTTAGATCAGCAGGCGGACGATCAAG 1486
      |:::   :::      ::::::|||::::::                 |
  407 lAlaGlyAsnAlaThrAsnThrIleIleAsn.................G 418

 1487 GCAAAAAACAAGCCTTTAGTGAAATCGGCTTGGTCAGCGGC......... 1527
      ||      |||   :::::::::   |||::::::::::|||
  418 lyGlyThrGlnAsnIleAsnAsnTyrGlyIleAlaThrGlyThrAsnIle 434

 1528 ...AGGGGGACGGTGCAACTGAATGCCGATAATCAGTTCAACCCCGACAA 1574
          ||||||   ::::::::::::   ::::::   :::
  435 AsnSerGlyThrGlnAsnIleLysSerGlyGlyLysAlaAspThrThrIl 451

 1575 ACTCTATTTCGGCTTTCGC.............................G 1594
       :::      |||   |||                             |
  451 eIleSerSerGlySerArgGlnValValGluLysAspGlyThrAlaIleG 468

 1595 GCGGACGTTTGGATTTGAACGGGCATTCGCTTTCGTTC............ 1632
      ||::::::::::::   :::|||      :::   :::
  468 lySerAsnIleSerAlaGlyGlySerLeuIleValTyrThrGlyGlyIle 484
```

```
1633 ...CACCGCATTCAAAATACCGATGAAGGGGCGATGATTGTCAAC..... 1674
       |||   :::::::::          :::|||:::::::::|||
 485 AlaHisGlyValAsnGlnGluThrGlySerAlaLeuValAlaAsnThrGl 501

1675 ...............CACAATCAAGACAAAGAATCCACCGTTACCATTA 1708
                     ::::::   :::|||   |||      |||||||
 501 yAlaGlyThrAspIleGluGlyTyrAsnLysLeuSerHisPheThrIleT 518

1709 CAGGCAATAAA.............................GATATTACT 1728
     |||||::::::                            ::::::|||
 518 hrGlyGlyGluAlaAsnTyrValValLeuGluAsnThrGlyGluLeuThr 534

1729 ............ACAACCGGCAATAACAACAACTTGGATAGCAAAAAAGA 1766
                |||:::::::::|||:::::::::|||:::      ::
 535 ValValAlaLysThrSerAlaLysAsnThrThrIleAspThrGlyGlyLy 551

1767 AATTGCCTACAACGGTTGGTTTGGCGAGAAAGATGCAACCAAAACGAACG 1816
     ::::                 :::|||:::|||         :::
 551 sLeuIleVal..............GlnLysGluAlaLysThrAspSerT 563

1817 GGCGGCTCAAT.................................CTG 1830
       |||||||||
 563 hrArgLeuAsnAsnGlyGlyValLeuGluValGlnAspGlyGlyGluAla 579

1831 AATTACCAACCGGAAGAAGCGGATCGCACTTTACTGCTTTCC........ 1872
     ::::::      :::::::::      :::|||:::   |||
 580 LysHisValGluGlnGlnSerGlyGlyAlaLeuIleAlaSerThrThrSe 596

1873 ............GGCGGAACAAATTTAAACGGCAAT............. 1896
                  |||||||||      |||:::
 596 rGlyThrLeuIleGluGlyThrAsnSerTyrGlyAspAlaPheTyrIleA 613

1896 ..................................................... 1896

 613 rgAsnSerGluAlaLysAsnValValLeuGluAsnAlaGlySerLeuThr 629

1897 ............................ATCACGCAAACAAACGGCAA 1916
                                 |||   :::::|||||||||
 630 ValValThrGlySerArgAlaValAspThrIleIleAsnAlaAsnGlyLy 646

1917 ACTGTTTTTCAGCGGCAGACCGACACCGCACGCCTACAATCATTTAGGAA 1966
     |:::        |||:::  :::      :::   |||      |||:
 646 sMetAspValTyrGlyLysAspValGlyThrValLeuAsnSerAlaGlyT 663

1967 GC..................................................... 1968
     ::
 663 hrGlnThrIleTyrAlaSerAlaThrSerAspLysAlaAsnIleLysGly 679

1969 ................GGGTGGTCAAAAATGGAAGGTATCCCACAAGG 2000
                     |||   :::         :::|||   :::||
 680 GlyLysGlnThrValTyrGlyLeuAlaThrGluAlaAsnIleGluSerGl 696

2001 AGAAATCGTGTGGGACAACGATTGGATCGACCGCACATTTAAAGCGGAAA 2050
     ||||   :::   |||:::       ::::::|||
 696 yGluGlnIleValAspGlyGlySerThrGluLysThr.............. 708

2051 ACTTCCATATTCAGGGCGGA............................. 2070
           ||||||:::||||||
 709 .....HisIleAsnGlyGlyThrGlnThrValGlnAsnTyrGlyLysAla 723

2071 ........................CAAGCGGTGGTTTCCCGCAATGT 2093
                             |||   :::::::::::::::::
 724 IleAsnThrAspIleValSerGlyLeuGlnGlnIleMetAlaAsnGlyTh 740

2094 TGCCAAA.........GTGGAAGGCGATTGGCATTTAAGCAATCACGCCC 2134
     :|||:::         :::::::|||      ::::::   |||::::::
 740 rAlaGluGlySerIleIleAsnGlyGlySerGlnValValAsnGluGlyG 757

2135 AAGCAGTTTTCGGTGTCGCACCGCATCAAAGCCACACAATCTGTACACGT 2184
           :::   :::   :::   :::::::::   |||:::   :::|||
 757 lyLeuAlaGluAsnSerValLeuAsnAspGlyGlyThrLeuAspValArg 773
```

```
2185 ......TCGGACTGGACGGGTCTG........................ 2202
           ::::::   ||||||:::
 774 GluLysGlySerAlaThrGlyIleGlnGlnSerSerGlnGlyAlaLeuVa 790

2203 .......ACAAGTTGTACCGAAAAAACCATTACCGACGATAAAGTGATTG 2245
            |||   :::|||:::   |||   |||            :::|
 790 lAlaThrThrArgAlaThrArgValThrGlyThrArgAlaAspGlyValA 807

2246 CTTCATTGAGCAAGACCGACATCAGAGGCAATGTCAGCCTTGCCGATCAC 2295
     ||          :::          :::|||:::   |||||||:::
 807 laPheSerIleGluGlnGlyAlaAlaAsnAsnIleLeuLeuAlaAsn... 822

2296 GCTCATTTAAATCTCACAGGACTTGCCACACTCAACGGCAATCTTAGTGC 2345
                          |||:::   |||:::::::::::::   |||::
 823 ..............GlyGlyValLeuThrValGluSerAspThrSerSe 834

2346 A.....................GGCGGAGACACGCACTATACGGTTACGC 2374
     :                      ||||||      :::      |||
 834 rAspLysThrGlnValAsnMetGlyGly......ArgGluIleValLysT 849

2375 GCAACGCCACCCAAAACGGCAACCTCAGCCTCGTGGGCAATGCCCAA... 2421
       :::||||||   :::|||:::   :::|||:::|||:::   |||
 849 hrLysAlaThrAlaThrGlyThr...ThrLeuThrGlyGlyGluGlnIle 864

2422 ......GCAACATTTAATCAAGCCACATTAAAC......GGCAACACATC 2459
           ::::::   |||:::::::|||:::|||      |||      ::
 865 ValGluGlyValAlaAsnGluThrThrIleAsnAspGlyGlyIleGlnTh 881

2460 GGCTTCGGACAATGCTTCATTTAATCTAAGCAACAACGCCGTACAAAACG 2509
     ::::|||   |||::::::      ::::::::::::      ::::::|
 881 rValSerAlaAsnGlyGlu...AlaIleLysThrLysIleAsnGluGlyG 897

2510 GCAGTCTGACGCTTTCCGACAACGCTAAGGCAAAC......GTAAGCCAT 2553
     ||:::||||||:::::|||||||:::|||||||:::      |||::::::
 897 lyThrLeuThrValAsnAspAsnGlyLysAlaThrAspIleValGlnAsn 913

2554 TCCGCACTCAACGGCAATGTCTCCCTAGCCGATAAGGCAGTATTCCATTT 2603
     |||:::          :::::::|||   |||:::          :::
 914 SerGlyAlaAlaLeuGlnThrSerThrAlaAsnGlyIleGluIleSerGl 930

2604 TGAAAACAGCCGCTTTACCGGAAAAATCAGCGGCGGCAAGGATACGGCAT 2653
        ::::::      |||   :::||||||||||:::      |||   :
 930 yThrHisGlnTyrGlyThrPheSerIleSerGlyAsnLeuAlaThrAsnM 947

2654 TACACTTAAAAGACAGC.............................. 2670
     ::   |||::::::::::
 947 etLeuLeuGluAsnGlyGlyAsnLeuLeuValLeuAlaGlyThrGluAla 963

2671 ...GAATGGACGCTGCCGTCGGGCACGGAATTAGGCAATTTAAACCTTGA 2717
        :::   |||:::   :::|||      :::   |||||||:::   ||
 964 ArgAspSerThrValGlyLysGlyGlyAlaMetGlnAsnLeuGlyGlnAs 980

2718 CAACGCCACCATTACACTCAATTCCGCCTATCGACAC.............. 2754
     |:::||||||      :::|||||||:::   ::::::
 980 pSerAlaThr...LysValAsnSerGlyGlyGlnTyrThrLeuGlyArgS 996

2755 .....................................GATGCGGCA 2763
                                         :::::|||
 996 erLysAspGluPheGlnAlaLeuAlaArgAlaGluAspLeuGlnValAla 1012

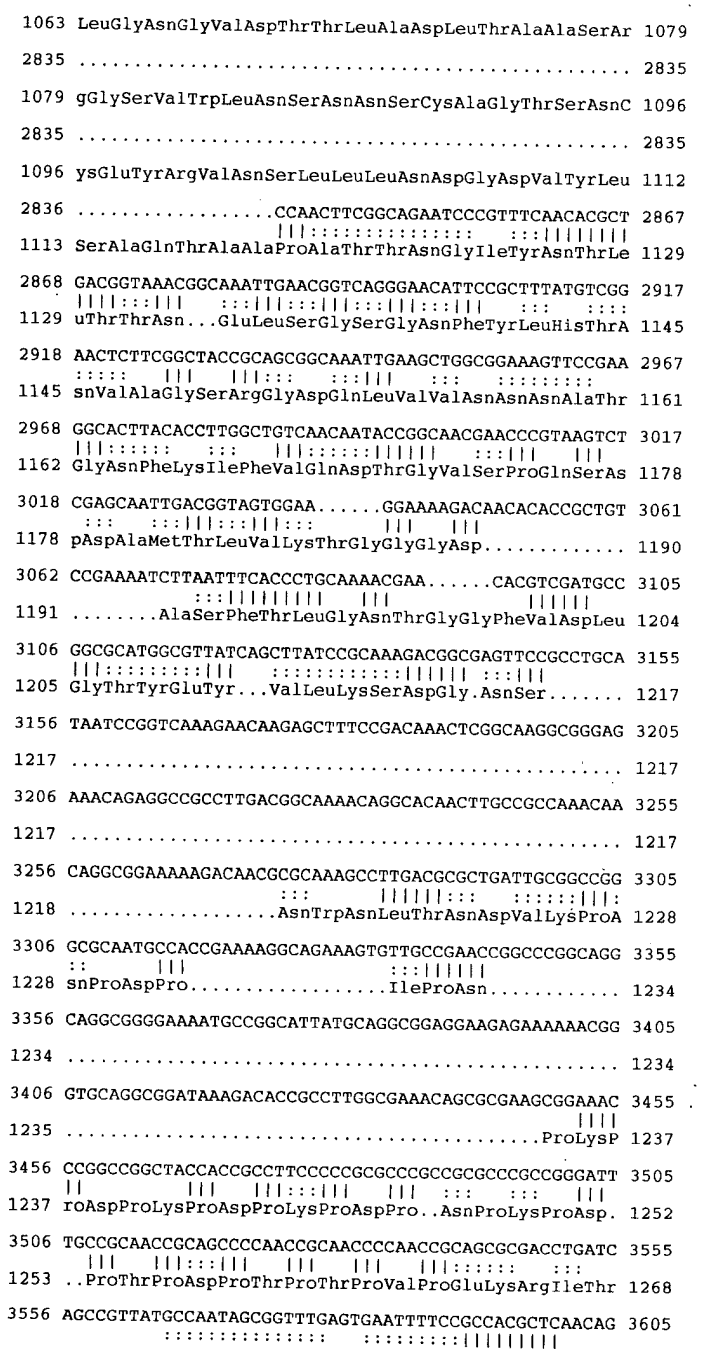
2764 GGC............GCGCAAACCGGCAGTGCGGCAGATGCGCCGCGCCG 2801
     |||            :::   :::|||:::   |||||||||
1013 GlyGlyThrAlaIleValTyrAlaGlyThrLeuAlaAspAlaSerValSe 1029

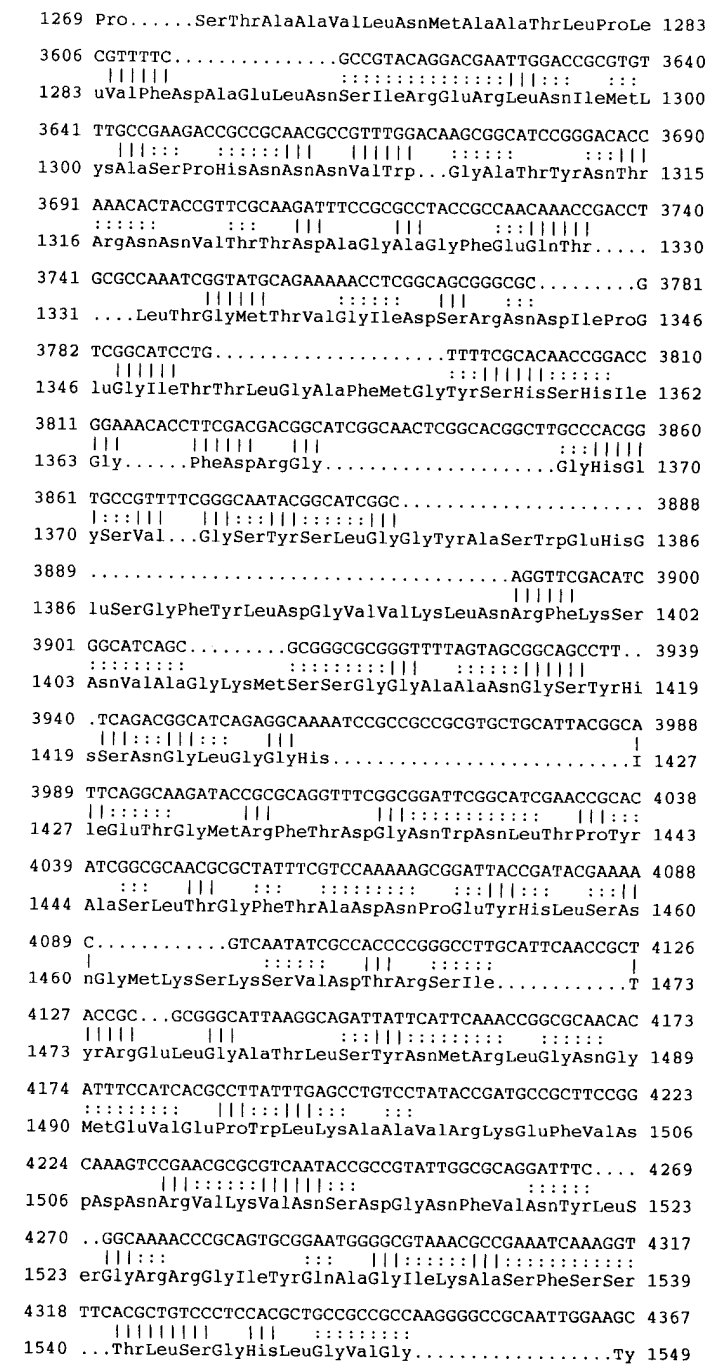
2802 CCGTTCGCGCCGTTCCCTATTATCCGTTACGCCG................ 2835
        :::      ||||||      :::|||||||
1029 rGlyAlaThrGlySerLeuSerLeuMetThrProArgAspAsnValThrP 1046

2835 ..................................................... 2835

1046 roValLysLeuGluGlyAlaValArgIleThrAspSerAlaThrLeuThr 1062

2835 ..................................................... 2835
```


1063 LeuGlyAsnGlyValAspThrThrLeuAlaAspLeuThrAlaAlaSerAr 1079
2835 2835
1079 gGlySerValTrpLeuAsnSerAsnAsnSerCysAlaGlyThrSerAsnC 1096
2835 2835
1096 ysGluTyrArgValAsnSerLeuLeuLeuAsnAspGlyAspValTyrLeu 1112
2836 .....CCAACTTCGGCAGAATCCCGTTTCAACACGCT 2867
1113 SerAlaGlnThrAlaAlaProAlaThrThrAsnGlyIleTyrAsnThrLe 1129
2868 GACGGTAAACGGCAAATTGAACGGTCAGGGAACATTCCGCTTTATGTCGG 2917
1129 uThrThrAsn...GluLeuSerGlySerGlyAsnPheTyrLeuHisThrA 1145
2918 AACTCTTCGGCTACCGCAGCGGCAAATTGAAGCTGGCGGAAAGTTCCGAA 2967
1145 snValAlaGlySerArgGlyAspGlnLeuValValAsnAsnAsnAlaThr 1161
2968 GGCACTTACACCTTGGCTGTCAACAATACCGGCAACGAACCCGTAAGTCT 3017
1162 GlyAsnPheLysIlePheValGlnAspThrGlyValSerProGlnSerAs 1178
3018 CGAGCAATTGACGGTAGTGGAA......GGAAAAGACAACACACCGCTGT 3061
1178 pAspAlaMetThrLeuValLysThrGlyGlyGlyAsp..... 1190
3062 CCGAAAATCTTAATTTCACCCTGCAAAACGAA......CACGTCGATGCC 3105
1191 .....AlaSerPheThrLeuGlyAsnThrGlyGlyPheValAspLeu 1204
3106 GGCGCATGGCGTTATCAGCTTATCCGCAAAGACGGCGAGTTCCGCCTGCA 3155
1205 GlyThrTyrGluTyr...ValLeuLysSerAspGly.AsnSer..... 1217
3156 TAATCCGGTCAAAGAACAAGAGCTTTCCGACAAACTCGGCAAGGCGGGAG 3205
1217 1217
3206 AAACAGAGGCCGCCTTGACGGCAAAACAGGCACAACTTGCCGCCAAACAA 3255
1217 1217
3256 CAGGCGGAAAAAGACAACGCGCAAAGCCTTGACGCGCTGATTGCGGCCGG 3305
1218 .....AsnTrpAsnLeuThrAsnAspValLysProA 1228
3306 GCGCAATGCCACCGAAAAGGCAGAAAGTGTTGCCGAACCGGCCCGGCAGG 3355
1228 snProAspPro.....IleProAsn..... 1234
3356 CAGGCGGGGAAAATGCCGGCATTATGCAGGCGGAGGAAGAGAAAAAACGG 3405
1234 1234
3406 GTGCAGGCGGATAAAGACACCGCCTTGGCGAAACAGCGCGAAGCGGAAAC 3455
1235 .....ProLysP 1237
3456 CCGGCCGGCTACCACCGCCTTCCCCCGCGCCCGCCGCGCCCGCCGGGATT 3505
1237 roAspProLysProAspProLysProAspPro..AsnProLysProAsp. 1252
3506 TGCCGCAACCGCAGCCCCAACCGCAACCCCAACCGCAGCGCGACCTGATC 3555
1253 ..ProThrProAspProThrProThrProValProGluLysArgIleThr 1268
3556 AGCCGTTATGCCAATAGCGGTTTGAGTGAATTTTCCGCCACGCTCAACAG 3605


1269 Pro.....SerThrAlaAlaValLeuAsnMetAlaAlaThrLeuProLe 1283
3606 CGTTTTC.....GCCGTACAGGACGAATTGGACCGCGTGT 3640
1283 uValPheAspAlaGluLeuAsnSerIleArgGluArgLeuAsnIleMetL 1300
3641 TTGCCGAAGACCGCCGCAACGCCGTTTGGACAAGCGGCATCCGGGACACC 3690
1300 ysAlaSerProHisAsnAsnAsnValTrp...GlyAlaThrTyrAsnThr 1315
3691 AAACACTACCGTTCGCAAGATTTCCGCGCCTACCGCCAACAAACCGACCT 3740
1316 ArgAsnAsnValThrThrAspAlaGlyAlaGlyPheGluGlnThr..... 1330
3741 GCGCCAAATCGGTATGCAGAAAAACCTCGGCAGCGGGCGC.....G 3781
1331 ....LeuThrGlyMetThrValGlyIleAspSerArgAsnAspIleProG 1346
3782 TCGGCATCCTG.....TTTTCGCACAACCGGACC 3810
1346 luGlyIleThrThrLeuGlyAlaPheMetGlyTyrSerHisSerHisIle 1362
3811 GGAAACACCTTCGACGACGGCATCGGCAACTCGGCACGGCTTGCCCACGG 3860
1363 Gly.....PheAspArgGly.....GlyHisGl 1370
3861 TGCCGTTTTCGGGCAATACGGCATCGGC..... 3888
1370 ySerVal...GlySerTyrSerLeuGlyGlyTyrAlaSerTrpGluHisG 1386
3889 .....AGGTTCGACATC 3900
1386 luSerGlyPheTyrLeuAspGlyValValLysLeuAsnArgPheLysSer 1402
3901 GGCATCAGC.....GCGGGCGCGGGTTTTAGTAGCGGCAGCCTT.. 3939
1403 AsnValAlaGlyLysMetSerSerGlyGlyAlaAlaAsnGlySerTyrHi 1419
3940 .TCAGACGGCATCAGAGGCAAAATCCGCCGCCGCGTGCTGCATTACGGCA 3988
1419 sSerAsnGlyLeuGlyGlyHis.....I 1427
3989 TTCAGGCAAGATACCGCGCAGGTTTCGGCGGATTCGGCATCGAACCGCAC 4038
1427 leGluThrGlyMetArgPheThrAspGlyAsnTrpAsnLeuThrProTyr 1443
4039 ATCGGCGCAACGCGCTATTTCGTCCAAAAAGCGGATTACCGATACGAAAA 4088
1444 AlaSerLeuThrGlyPheThrAlaAspAsnProGluTyrHisLeuSerAs 1460
4089 C.....GTCAATATCGCCACCCCGGGCCTTGCATTCAACCGCT 4126
1460 nGlyMetLysSerLysSerValAspThrArgSerIle.....T 1473
4127 ACCGC...GCGGGCATTAAGGCAGATTATTCATTCAAACCGGCGCAACAC 4173
1473 yrArgGluLeuGlyAlaThrLeuSerTyrAsnMetArgLeuGlyAsnGly 1489
4174 ATTTCCATCACGCCTTATTTGAGCCTGTCCTATACCGATGCCGCTTCCGG 4223
1490 MetGluValGluProTrpLeuLysAlaAlaValArgLysGluPheValAs 1506
4224 CAAAGTCCGAACGCGCGTCAATACCGCCGTATTGGCGCAGGATTTC.... 4269
1506 pAspAsnArgValLysValAsnSerAspGlyAsnPheValAsnTyrLeuS 1523
4270 ..GGCAAAACCCGCAGTGCGGAATGGGGCGTAAACGCCGAAATCAAAGGT 4317
1523 erGlyArgArgGlyIleTyrGlnAlaGlyIleLysAlaSerPheSerSer 1539
4318 TTCACGCTGTCCCTCCACGCTGCCGCCGCCAAGGGGCCGCAATTGGAAGC 4367
1540 ...ThrLeuSerGlyHisLeuGlyValGly.....Ty 1549

```
    4368 GCAGCACAGCGCGGGCATCAAA.................TTAGGCTACC 4399
         :::||||||||||||||::::::                  ||||   :
    1549 rSerHisSerAlaGlyValGluSerProTrpAsnAlaValAlaGlyValA 1566

    4400 GCTGG 4404
         ::|||
    1566 snTrp 1567

seq_name: pir2:B90674

seq_documentation_block:
AidA-I adhesin-like protein [imported] - Escherichia coli (strain O157:H7, substrain RIM
C;Species: Escherichia coli
C;Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 18-Jul-2001
C;Accession: B90674
R;Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.
gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.
DNA Res. 8, 11-22, 2001
A;Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and genc
A;Reference number: A99629; MUID:21156231; PMID:11258796
A;Accession: B90674
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1327 <HAY>
A;Cross-references: GB:BA000007; PIDN:BAB33785.1; PID:g13359819; GSPDB:GN00154
A;Experimental source: strain O157:H7, substrain RIMD 0509952
C;Genetics:
A;Gene: ECs0362


alignment_scores:
            Quality:  348.00              Length:   1373
              Ratio:  0.509                Gaps:      72
 Percent Similarity:  49.818   Percent Identity: 20.612

alignment_block:
 US-09-303-518D-653 x B90674   ..

 Align seg 1/1  to: B90674  from: 1  to: 1327

    1024 ATCGATGCCAAACATAAACACTATTCTCTACCTTATAGATTAAAAACACG 1073
         ||||||         ::::::   |||:::   |||||||:::   :::
      59 IleAspPheGlyLysGluAsnGlnSerIleAspTyrArgIleThrValTh 75

    1074 AACCGTTCAATTGTTTAATGTTTCTTTATCCGAGACAGCAAGAGAACCTG 1123
                  :::   ::::::   ::::::|||:::|||      :
      75 rAspAsnAlaAsnLeuValIleAsnAlaThrAspThrSerArgProArgL 92

    1124 TTTATCATGCTGCAGGTGGG.........GTCAACAGTTATCGACCCAGA 1164
         ::      |||:::||||||         :::::::::   :::
      92 euThrLeuAlaSerGlyGlyGlyLeuAspIleThrGlyGlyLysValThr 108

    1165 CTGAATAATGGAGAAAATATTTCCTTTATTGACAAAGGAAAAGGTGAATT 1214
         :::|||:::         ::::::|||:::   ||||||   |||   ||
     109 IleAsnGly......ProLeuAsnPheLeuLeuLysGlyThrGlyPheLe 123

    1215 GATACTTACCAGCAACATCAACCAAGGCGCGGGCGGTTTGTATTTTGAG. 1263
         |          |||:::::::::::::::::   ||||||   :::
     123 u...........AsnValSerAsnAlaGlySerGluLeuTyrAlaAspA 136

    1264 .....................................GGTAATTTTACG 1275
                                             |||   |||:::
     136 spLeuTyrGluSerAsnSerGlyMetArgHisAspArgGlyTyrPheAsn 152

    1276 GTCTCGCCTAAAAACAACGAAACGTGGCAAGGCGCGGGCGTTCATATCAG 1325
         ||||||                  :::|||:::   :::|||:::::
     153 ValSer...................AsnGlyGlyLysIleHisValLy 162

    1326 TGATGGCAGTACCGTTACTTGGAAAGTAAACGGCGTGGCAAACGACCGCC 1375
         :      |||   :::|||:::      ::::::|||:::::::::
     162 sGlyThrSerArgLeuThrTyrLeuGlnGlyAsnValSerGlyGluGly. 178

    1376 TGTCCAAAATCGGCAAAGGCACGCTGCTGGTTCAAGCCAAAGGGGAAAAC 1425
          |||:::::::::::::   |||:::::::::   :::   |||:::
     179 ..SerGlnValAsnSerGluThrPhePheMetGlyValTyrGlySerTyr 194

    1426 CAAGGCTCG......GTCAGCGTGGGCGACGGTAAAGTCATC........ 1461
          |||:::      :::||||||::::::|||      :::
     195 GlyGlyAsnGlnTyrLeuSerValAsnAsnGlyGlyGluValAsnAlaAr 211

    1462 ......................TTAGATCAGCAGGCGGACGATCAAGGCA 1489
                                 ||||||   :::|||
     211 gLysGlnIleSerLeuGlyTyrTyrAspGlnValSerAsp.........T 225

    1490 AAAAACAAGCCTTTAGTGAAATCGGCTTGGTCAGCGGCAGGGGGACGGTG 1539
               |||   ||||||   |||   :::|||:::      |||:::
     225 hrThrLeuAlaValSerGluGlyGlyLysIleSerAla...ProThrIle 240

    1540 CAACTGAATGCCGATAATCAGTTCAACCCCGACAAACTCTATTTCGGCTT 1589
         :::|||:::::::::::::::                 |||   :::|||
     241 SerLeuSerThrAsnSerGlu..............LeuAlaLeuGlyAl 252

    1590 TCGCGGCGGACGTTTGGATTTGAACGGGCATTCGCTTTCGTTCCACCGCA 1639
          :::   |||              :::      ::::::   :::::|
     252 aGlnGluGlySerAlaAlaLysAlaAlaGlyIleIleAspAlaGluLysI 269

    1640 TTCAA............AATACCGATGAAGGGGCGATGATTGTCAACCAC 1677
         ||:::            :::|||:::|||      :::   :::||||||
     269 leGluPheValTrpAlaLysThrSerGluLysLysIleThrLeuAsnHis 285

    1678 AATCAAGACAAAGAATCCACCGTTACC.........ATTACAGGCAATAA 1718
         :::   ||||||::::::|||::::::         ::::::|||:::::
     286 Thr...AspLysAspAlaThrIleSerAlaAspIleValSerGlySerGl 301

    1719 AGATATT.......................ACTACA......ACCGGCA 1738
         :   :::                       ||||||      ||||||:
     301 uGlyLeuGlyTyrIleAsnAlaLeuAsnGlyThrThrTyrLeuThrGlyA 318

    1739 ATAACAACAACTTGGATAGCAAAAAAGAAATTGCCTACAACGGTTGGTTT 1788
         ::|||:::   ::::::::::|||   :::|||      ||||||   :::
     318 spAsnSerAlaPheSerGlyLysValLysIleGluGlnAsnGlyAlaLeu 334

    1789 GGCGAGAAAGATGCAACCAAAACGAACGGGCGGCTCAATCTGAATTACCA 1838
         |||            |||::::::   |||      :::::::|||   ::
     335 Gly.........IleThrGlnAsnIleGlyThrAlaGluIleAsnAsnAr 348

    1839 ACCGGAA.........GAAGCGGATCGCACTTTACTGCTTTCCGGCGGAA 1879
         :   :::         :::||||||   ::::::   :::::::::
     348 gGlyLysLeuHisLeuLysAlaAspAspSerMetThrPheAlaAsnLysI 365

    1880 CAAATTTAAACGGCAATATCACGCAAACAAACGGCAAACTGTTTTTCAGC 1929
           :::   ||||||:::|||:::      :::|||   :::   ::::::
     365 leSerGlyAsnGlyThrIleSerIleAspSerGlyThrValGluLeuThr 381

    1930 GGCAGACCGACACCGCACGCCTACAATCATTTAGGAAGCGGGTGG..... 1974
         |||:::                  |||:::      ||||||:::
     382 GlyAsn..................AsnTyrAlaPheSerGlyTyrIleAs 392

    1975 ..........................TCAAAAATGGAAGGTATCCCAC 1996
                                    |||:::   ::::::|||   :
     392 pValAlaSerGlyAlaValAlaValIleSerGluAspLysAsnIleGlyA 409

    1997 AAGGAGAAATCGTGTGGGAC.......................AACGAT 2022
         :::::|||:::      |||                       :::|||
     409 rgAlaGluLeuAspValAspGlyLysLeuGlnIleAsnAlaAsnLysAsp 425

    2023 TGGATC...GACCGCACATTTAAAGCGGAAAACTTCCATATTCAGGGCGG 2069
         |||:::   |||:::   ::::::                        ||
     426 TrpValPheAspAsnAspLeuGlu.......................Gl 434

    2070 ACAAGCGGTGGTTTCCCGCAATGTTGCCAAAGTGGAA............. 2106
         |:::::::::|||:::   |||:::::::::   |||
     434 yArgGlyIleValGluIleAsnMetGlyAsnHisGluPheSerPheAspG 451

    2107 ..........GGCGATTGG........................... 2115
```

```
                  ||||||
 451 luPheAlaTyrThrAspTrpPheGlnGlySerLeuAlaPheGlnAsnThr 467

2116 ......CATTTAAGCAATCACGCCCAAGCAGTT......TTCGGTGTCGC 2153
          :::|||:::::::::|||:::   :::         |||:::::
 468 ThrPheAsnLeuGluLysAsnAlaGluPheLeuGlnLysGlyGlyIleTh 484

2154 ACCGCATCAA.........................AGCCACACAATCT 2176
      :   |||                          :::|||:::|||
 484 rAlaGlyGlnGlySerLeuValThrValGlyLysGlyAlaHisSerIleS 501

2177 GTACACGTTCG.................GACTGGACGGGTCTGACAAGT 2208
       |||   :::                 |||:::   :::||||||:::
 501 erThrLeuGlyPheSerGlyGlyThrValAspPheGlyAlaLeuThrAla 517

2209 .........TGTACCGAAAAAACCATTACCGACGATAAAGTG........ 2241
             ||||||   |||::::::   :::|||:::
 518 GlyAlaGlnMetThrGluGlyThrValAsnValSerLysThrLeuAspLe 534

2242 ............ATTGCTTCATTGAGCAAGACCGACATCAGAGGCAATG 2278
             :::   :::::::|||   :::|||:::      :::|
 534 uArgGlyGluGlyValIleGlnValSerAspSerAspValValArgServ 551

2279 TCAGCCTTGCCGATCACGCTCATTTAAATCTCACA............GGA 2316
     |||||          :::   |||:::||||||           |||
 551 alSerArgAspIleAspSerAlaLeuSerLeuThrGluValAspAspGly 567

2317 CTTGCCACACTCAAC................................ 2331
       :::|||::::::
 568 AsnSerThrIleLysLeuValAspAlaGlnGlyAlaGluValLeuGlyAs 584

2332 ....GGCAATCTTAGTGCA............GGCGGAGACACGCACTATA 2365
         |||||||||:::                |||           :
 584 pAlaGlyAsnLeuGlnLeuGlnAspLysAsnGlyGlnIleLeuSerSerS 601

2366 CGGTTACGCGCAACGCCACCCAAAACGGCAACCTCAGCCTCGTGGGCAAT 2415
     :::::   |||:::      ||||||||||:::   :::   ||||||
 601 erAlaGlnArgAspIleGlnGlnAsnGlyGlnLysAlaAlaValGly... 616

2416 GCCCAAGCAACATTTAATCAAGCCACATTAAACGGCAACACATCGGCTTC 2465
           |||:::::             :::|||   ::::::   ::
 617 .........ThrTyrAspTyrArgLeuThrSerGlyValAsnAsnAspGl 630

2466 GGACAATGCTTCATTTAATCTAAGCAAC.........AACGCCGTACAAA 2506
     :         :::::::::|||:::::          :::|||:::::::
 630 yLeuTyrIleGlyTyrGlyLeuThrGlnLeuAspLeuHisAlaThrAspS 647

2507 ACGGCAGTCTGACGCTTTCCGACAACGCTAAGGCAAACGTAAGCCATTCC 2556
     ::   :::|||:::|||||||:::|||:::|||         |||:::::
 647 erAspAlaLeuValLeuSerSerAsnGlyLys.........SerGluAsn 660

2557 GCACTCAACGGCAATGTCTCCCTAGCCGATAAGGCAGTATTCCATTTTGA 2606
     |||   :::   ::::::::::::::::   ::::::   :::   |||::
 661 AlaAlaAspLeuSerAlaLysIleThrGlySerGlyAspLeuAlaPheSe 677

2607 AAACAGCCGCTTTACCGGAAAAATCAGCGGCGGCAAGGATACGGCATTAC 2656
     ::::::::::                 |||:::   ::::::|||
 677 rSerGlnLys...................GlyGlnThrValSerLeuS 687

2657 ACTTAAAAGACAGCGAATGGACGCTGCCGTCGGGCACGGAATTAGGCAAT 2706
          ||||||::::::::::|||     :::      :::   |||:::
 687 erAsnLysAspAsnAspTyrThrGlyValThrAspLeuArgSerGlyThr 703

2707 TTAAACCTTGACAACGCCACCATTACACTCAATTCCGCCTATCGACACGA 2756
     |||   |||:::|||   ::::::      |||:::         |||::
 704 LeuLeuLeuAsnAsnAspAsnValLeuGlyAsnThr.........HisGl 717

2757 TGCGGCAGGCGCGCAAACCGGCAGTGCGGCAGATGCGCCGCGCCGCCGTT 2806
     :
 717 u..................................................  717

2807 CGCGCCGTTCCCTATTATCCGTTACGCCGCCAACTTCGGCAGAATCCCGT 2856
                |||   ::::::      |||:::   :::   :::
```

```
 718 ..............LeuArgLeuAlaAlaGluThrGluLeuAspMetAsn 729

2857 TTCAACACGCTGACGGTAAACGGCAAATTGAACGGTCAGGGAACATTCCG 2906
       ::::::   ||||||   |||   |||||||||:::::
 730 GlyHisSerGlnThrVal...GlyThrLeuAsnGlySerAla........ 742

2907 CTTTATGTCGGAACTCTTCGGCTACCGCAGCGGCAAATTGAAGCTGGCGG 2956
           :::::::|||::::::   :::::::|||:::|||   :::::::
 743 .......AspSerLeuLeuSerLeuAsnGlyGlySerLeuThrValThrA 757

2957 AAAGTTCCGAAGGCACTTACACCTTGGCTGTCAACAATACCGGCAACGAA 3006
     ::::::::   :::|||   :::|||            ||||||:::
 757 snGlyGlyThrSerThrGlySerLeu............ThrGlySerGly 769

3007 CCCGTAAGTCTCGAGCAATTGACGGTAGTGGAAGGAAAAGACAACACACC 3056
       ::::::::::::       |||:::      :::   ||||||:::
 770 GluLeuAsnIleGlnGlyGlyThrLeuAspIleAlaGlyAspAsnSerAs 786

3057 GCTGTCCGAAAATCTTAATTTC..............ACCCTGCAAAACG 3091
      |||:::   |||:::|||:::               :::|||   ::::
 786 nLeuThrAlaAsnValAsnIleAlaAsnSerAlaAsnValLeuValSerH 803

3092 AACACGTCGATGCCGGCGCATGGCGTTATCAGCTTATCCGCAAAGACGGC 3141
     ::            |||:::   :::       :::::::::::|||
 803 isAlaGlnGlyLeuGlySerAlaAsn.........ValGluAsnAsnGly 816

3142 GAGTTCCGCCTGCATAATCCGGTCAAAGAACAA.................. 3174
       :::   |||:::|||   ::::::::::::
 817 ThrLeuAlaLeuAsnAsnSerAlaGluLysArgAlaAlaAlaSerValAs 833

3175 ....GAGCTTTCCGACAAACTCGGCAAGGCGGGAGAAACAGAGGCCGCCT 3220
           |||:::   :::|||   :::   |||          :::::::
 833 nTyrAlaLeuGlyGlyAsnLeuThrAsnAsnGlyThrLeuMetThrGlyM 850

3221 TGACGGCAAAACAGGCA......CAACTTGCCGCCAAACAACAGGCGGAA 3264
     ::::::::::::||||||          |||:::::|||   :::   :::
 850 etSerGlyGlnGlnAlaGlyAsnValLeuValValLysGlyAsnTyrHis 866

3265 AAAGACAACGCGCAAAGCCTT............................. 3285
       :::|||:::|||   :::
 867 GlyAsnAsnGlyGlnLeuValMetAsnThrValLeuAsnGlyAspAspSe 883

3286 .......GACGCGCTGATTGCGGCCGGGCGCAATGCCACCGAAAAGGCAG 3328
           |||   |||::::::   |||             :::   :::
 883 rValThrAspLysLeuValValGluGly............AspThrSerG 896

3329 AAAGTGTTGCCGAACCGGCCCGGCAGGCAGGCGGGGAAAATGCC...... 3372
       :::::::|||      :::::::::|||||||||   :::|||
 896 lyThrThrAlaValThrValAsnAsnAlaGlyGlyThrGlyAlaLysThr 912

3373 ......GGCATT............ATGCAGGCGGAGGAAGAGAAAAAACG 3404
           ||||||             :::::::::::::::|||   :::
 913 LeuAsnGlyIleGluLeuIleHisValAspGlyLysSerGluGlyGluPh 929

3405 GGTGCAGGCGGATAAAGACACCGCCTT....................... 3431
      |||||||||   :::   :::|||
 929 eValGlnAlaGlyArgIleValAlaGlyAlaTyrAspTyrThrLeuAlaA 946

3432 ..GGCGAAACAGCGCGAAGCGGAAAC..............CCGGCCGGC 3464
       |||:::   |||:::||||||||||                 :::
 946 rgGlyGlnGlyAlaAsnSerGlyAsnTrpTyrLeuThrSerGlySerAsp 962

3465 TACCACCGCCTTCCCCCGCGCCCG..............CCGCGCCCGCC 3499
           :::|||   |||:::|||               |||:::||| :
 963 SerProGluLeuGlnProGluProAspProMetProAsnProGluPro.A 979

3500 GGGATTTGCCGCAACCGCAGCCCCAACCGCAACCCCAACCGCAGCGCGAC 3549
     ::      |||:::|||:::|||:::|||   |||   |||       |||
 979 snProAsnProGluProAsnProAsnProThrProThrProGlyProAsp 995

3550 CTG..............ATCAGCCGTTATGCCAATAGCGGTTTGAGTGA 3584
     |||                 ::         |||:::|||   ::::::::
 996 LeuAsnValAspAsnAspLeuArgProGluAlaGlySerTyrIleAlaAs 1012
```

```
3585 ATTTTCCGCCACGCTCAACAGCGTTTTCGCC......GTACAGGACGAAT 3628
     :::::::|||:::   |||:::::|||:::      ::::::::::::|
1012 nLeuAlaAlaAla...AsnThrMetPheThrThrArgLeuHisGluArgL 1028

3629 TG.................GACCGCGTGTTTGCCGAAGACCGCCGCAAC 3660
     ||                 |||   |||   :::|||:::::::::::::
1028 euGlyAsnThrTyrTyrThrAspMetValThrGlyGluGlnLysGlnThr 1044

3661 GCCGTTTGGACAAGCGGCATCCGGGACACCAAACACTACCGTTCGCAAGA 3710
     ::::::|||           ::::::     |||   :::   :::||
1045 ThrMetTrpMetArg......HisGluGlyGlyHisAsnLysTrpArgAs 1059

3711 TTTCCGCGCCTACCGCCAACAAACCGACCTGCGCCAAATCGGTATGCAGA 3760
     |
1059 p..................................................  1059

3761 AAAACCTCGGCAGCGGGCGCGTCGGCATCCTGTTTTCGCACAACCGGACC 3810
           |||||||||:::::                    :::
1060 ........GlySerGlyGlnLeu.................LysThrGln 1067

3811 GGAAACACCTTCGACGACGGCATCGGCAACTCGGCACGGCTTGCCCACGG 3860
     :::|||   :::         :::|||:::::::::
1068 SerAsnArgTyrValLeuGlnLeuGlyGlyAspVal.............  1079

3861 TGCCGTTTTCGGGCAATACGGCATCGGCAGGTTCGACATCGGCATCAGCG 3910
      |||   :::::::|||   |||      |||:::   :::|||:::   |
1080 .AlaGlnTrpSerGlnAsnGlySerAspArgTrpHisValGlyValMetA 1096

3911 CGGGCGCGGGTTTTAGTAGCGGC.........AGCCTTTCAGACGGCATC 3951
     |||||   |||   |||::::::         |||         |||
1096 laGlyTyrGlyAsnSerAspSerLysThrIleSerSerArgThrGlyTyr 1112

3952 AGAGGCAAAATCCGCCGCCGCGTGCTGCATTACGGCATTCAGGCAAGATA 4001
     |||:::|||          :::            |||:::   |||   ::
1113 ArgAlaLysAlaSerValAsnGlyTyrSerThrGlyLeuTyrAlaThrTr 1129

4002 CCGCGCAGGTTTCGGCGGATTCGGCATCGAACCGCACATCGGCGCAACGC 4051
     :   |||                  |||   :::   ||||||
1129 pTyrAla..............AspAspGluSerArgAsnGlyAla.... 1139

4052 GCTATTTCGTCCAAAAAGCGGATTACCGATACGAAAACGTCAATATCGCC 4101
       |||:::   :::   |||:::|||   :::   :::   ::::::
1140 ..TyrLeuAspSerTrpAlaGlnTyrSerTrpPheAspAsnThrValLys 1155

4102 ACCCCGGGCCTTGCATTCAACCGCTACCGC..............GCGGG 4136
               |||      :::   |||:::                |||::
1156 GlyAspAspLeuGlnSerGluSerTyrLysSerLysGlyPheThrAlaSe 1172

4137 CATTAAGGCAGATTATTCATTCAAACCGGCGCAACACATTTCC....... 4179
     :::::::|||   |||:::   |||   |||:::      :::
1172 rLeuGluAlaGlyTyrLysHisLysLeuAlaGluPheAsnGlySerGlnG 1189

4180 ...................ATCACGCCTTATTTGAGCCTGTCCTATACC 4209
                         :::   |||       ::::::::::::
1189 lyThrArgAsnGluTrpTyrValGlnProGlnAlaGlnValThrTrpMet 1205

4210 GATGCC..................................................  4215
        :::
1206 GlyValLysAlaAspLysHisArgGluSerAsnGlyThrLeuValHisSe 1222

4216 ....GCTTCCGGCAAAGTCCGAACGCGCGTCAATACCGCCGTATTGGCGC 4261
         ::::::|||:::|||:::|||||::::::::::   :::      :
1222 rAsnGlyAspGlyAsnValGlnThrArgLeuGlyValLysThrTrpLeuL 1239

4262 AG..............GATTTCGGCAAAACCCGC.............. 4281
     ::              |||   ||||||:::|||
1239 ysSerHisHisLysMetAspAspGlyLysSerArgGluPheGlnProPhe 1255

4282 ...AGTGCGGAATGGGGCGTAAACGCCGAAATCAAAGGTTTCACGCTGTC 4328
        ::::::::::|||   |   |||:::      |||   |||:::   ||
1256 ValGluValAsnTrpLeuHisAsnSer......LysAspPheSerThrSe 1270

4329 CCTCCACGCTGCCGCCGCCAAGGGGCCGCAATTGGAAGCGCAGCACAGCG 4378
     |:::   ::::::::::::         |||      |||::::::   |
1270 rMetAspGlyValSerVal......ThrGlnAspGlyAlaArgAsnIleA 1285

4379 CGGGCATCAAATTAGGC 4395
     ||   ||||||   |||
1285 laGluIleLysThrGly 1290
```

```
seq_name: pir2:E85524

seq_documentation_block:
probable beta-barrel outes membrane protein Z0402 [imported] - Escherichia coli (stra
C;Species: Escherichia coli
C;Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 14-Sep-2001
C;Accession: E85524
R;Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; May
iller, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apoda
Nature 409, 529-533, 2001
A;Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
A;Reference number: A85480; MUID:21074935; PMID:11206551
A;Accession: E85524
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1349 <STO>
A;Cross-references: GB:AE005174; NID:g12513130; PIDN:AAG54657.1; GSPDB:GN00145; UWGP:
A;Experimental source: strain O157:H7, substrain EDL933
C;Genetics:
A;Gene: Z0402


alignment_scores:
            Quality:  348.00              Length:   1373
              Ratio:  0.509                 Gaps:     72
 Percent Similarity:  49.818   Percent Identity: 20.612

alignment_block:
 US-09-303-518D-653 x E85524   ..

 Align seg 1/1  to: E85524  from: 1  to: 1349

   1024 ATCGATGCCAAACATAAACACTATTCTCTACCTTATAGATTAAAAACACG 1073
        ||||||         ::::::   |||:::   |||||||:::   :::
     81 IleAspPheGlyLysGluAsnGlnSerIleAspTyrArgIleThrValTh 97

   1074 AACCGTTCAATTGTTTAATGTTTCTTTATCCGAGACAGCAAGAGAACCTG 1123
                     :::   :::::   :::::::|||:::|||      :
     97 rAspAsnAlaAsnLeuValIleAsnAlaThrAspThrSerArgProArgL 114

   1124 TTTATCATGCTGCAGGTGGG.........GTCAACAGTTATCGACCCAGA 1164
        ::      |||:::||||||         :::::::::   :::
    114 euThrLeuAlaSerGlyGlyGlyLeuAspIleThrGlyGlyLysValThr 130

   1165 CTGAATAATGGAGAAAATATTTCCTTTATTGACAAAGGAAAAGGTGAATT 1214
        :::|||:::         :::::::|||:::   ||||||   |||   ||
    131 IleAsnGly......ProLeuAsnPheLeuLeuLysGlyThrGlyPheLe 145

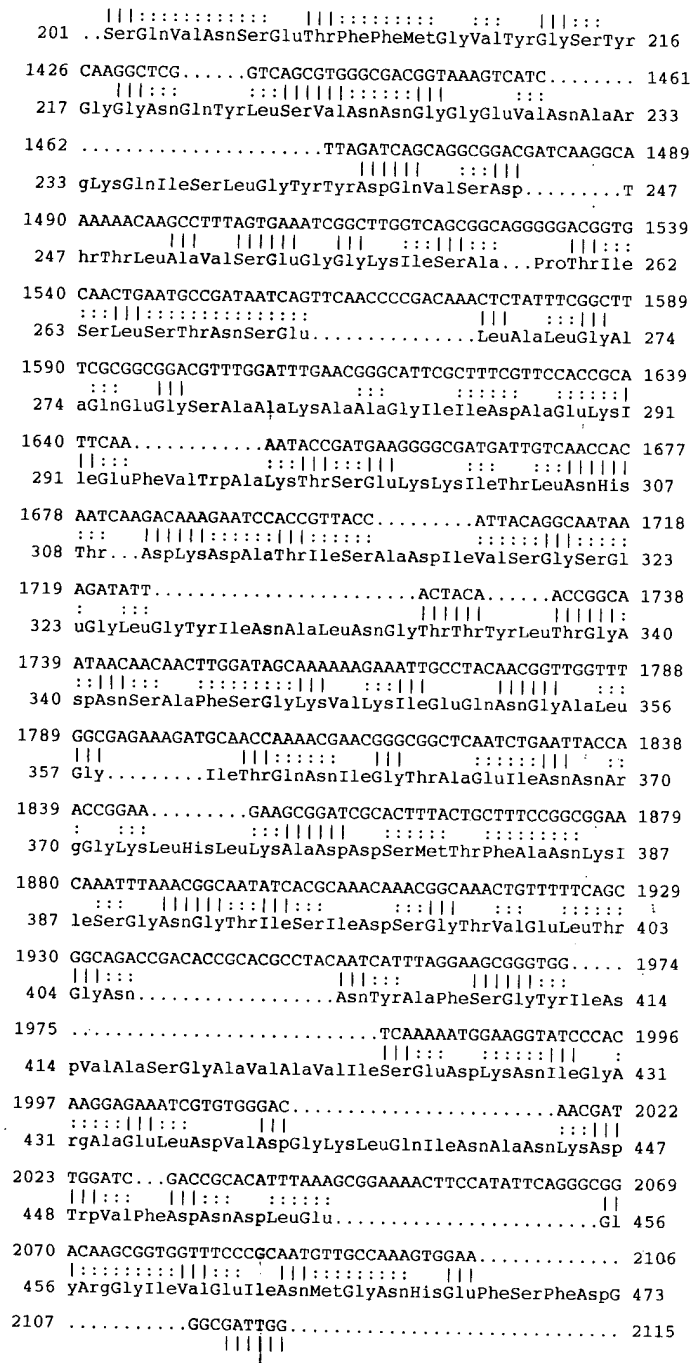
   1215 GATACTTACCAGCAACATCAACCAAGGCGCGGGCGGTTTGTATTTTGAG. 1263
        |            |||:::::::::::::::::   ||||||   :::
    145 u............AsnValSerAsnAlaGlySerGluLeuTyrAlaAspA 158

   1264 .....................................GGTAATTTTACG 1275
                                            |||   |||:::
    158 spLeuTyrGluSerAsnSerGlyMetArgHisAspArgGlyTyrPheAsn 174

   1276 GTCTCGCCTAAAAACAACGAAACGTGGCAAGGCGCGGGCGTTCATATCAG 1325
        ||||||                      :::|||:::   :::|||:::::
    175 ValSer...................AsnGlyGlyLysIleHisValLy 184

   1326 TGATGGCAGTACCGTTACTTGGAAAGTAAACGGCGTGGCAAACGACCGCC 1375
        :      |||   :::|||:::      ::::::|||:::::::::
    184 sGlyThrSerArgLeuThrTyrLeuGlnGlyAsnValSerGlyGluGly. 200

   1376 TGTCCAAAATCGGCAAAGGCACGCTGCTGGTTCAAGCCAAAGGGGAAAAC 1425
```


201 ..SerGlnValAsnSerGluThrPhePheMetGlyValTyrGlySerTyr 216
1426 CAAGGCTCG......GTCAGCGTGGGCGACGGTAAAGTCATC..... 1461
217 GlyGlyAsnGlnTyrLeuSerValAsnAsnGlyGlyGluValAsnAlaAr 233
1462 .....TTAGATCAGCAGGCGGACGATCAAGGCA 1489
233 gLysGlnIleSerLeuGlyTyrTyrAspGlnValSerAsp.....T 247
1490 AAAAACAAGCCTTTAGTGAAATCGGCTTGGTCAGCGGCAGGGGGACGGTG 1539
247 hrThrLeuAlaValSerGluGlyGlyLysIleSerAla...ProThrIle 262
1540 CAACTGAATGCCGATAATCAGTTCAACCCCGACAAACTCTATTTCGGCTT 1589
263 SerLeuSerThrAsnSerGlu.....LeuAlaLeuGlyAl 274
1590 TCGCGGCGGACGTTTGGATTTGAACGGGCATTCGCTTTCGTTCCACCGCA 1639
274 aGlnGluGlySerAlaAlaLysAlaAlaGlyIleIleAspAlaGluLysI 291
1640 TTCAA.....AATACCGATGAAGGGGCGATGATTGTCAACCAC 1677
291 leGluPheValTrpAlaLysThrSerGluLysLysIleThrLeuAsnHis 307
1678 AATCAAGACAAAGAATCCACCGTTACC.....ATTACAGGCAATAA 1718
308 Thr...AspLysAspAlaThrIleSerAlaAspIleValSerGlySerGl 323
1719 AGATATT.....ACTACA.....ACCGGCA 1738
323 uGlyLeuGlyTyrIleAsnAlaLeuAsnGlyThrThrTyrLeuThrGlyA 340
1739 ATAACAACAACTTGGATAGCAAAAAAGAAATTGCCTACAACGGTTGGTTT 1788
340 spAsnSerAlaPheSerGlyLysValLysIleGluGlnAsnGlyAlaLeu 356
1789 GGCGAGAAAGATGCAACCAAAACGAACGGGCGGCTCAATCTGAATTACCA 1838
357 Gly.....IleThrGlnAsnIleGlyThrAlaGluIleAsnAsnAr 370
1839 ACCGGAA.....GAAGCGGATCGCACTTTACTGCTTTCCGGCGGAA 1879
370 gGlyLysLeuHisLeuLysAlaAspAspSerMetThrPheAlaAsnLysI 387
1880 CAAATTTAAACGGCAATATCACGCAAACAAACGGCAAACTGTTTTTCAGC 1929
387 leSerGlyAsnGlyThrIleSerIleAspSerGlyThrValGluLeuThr 403
1930 GGCAGACCGACACCGCACGCCTACAATCATTTAGGAAGCGGGTGG..... 1974
404 GlyAsn.....AsnTyrAlaPheSerGlyTyrIleAs 414
1975 .....TCAAAAATGGAAGGTATCCCAC 1996
414 pValAlaSerGlyAlaValAlaValIleSerGluAspLysAsnIleGlyA 431
1997 AAGGAGAAATCGTGTGGGAC.....AACGAT 2022
431 rgAlaGluLeuAspValAspGlyLysLeuGlnIleAsnAlaAsnLysAsp 447
2023 TGGATC...GACCGCACATTTAAAGCGGAAAACTTCCATATTCAGGGCGG 2069
448 TrpValPheAspAsnAspLeuGlu.....Gl 456
2070 ACAAGCGGTGGTTTCCCGCAATGTTGCCAAAGTGGAA..... 2106
456 yArgGlyIleValGluIleAsnMetGlyAsnHisGluPheSerPheAspG 473
2107 .....GGCGATTGG..... 2115
473 luPheAlaTyrThrAspTrpPheGlnGlySerLeuAlaPheGlnAsnThr 489
2116 .....CATTTAAGCAATCACGCCCAAGCAGTT.....TTCGGTGTCGC 2153
490 ThrPheAsnLeuGluLysAsnAlaGluPheLeuGlnLysGlyGlyIleTh 506
2154 ACCGCATCAA.....AGCCACACAATCT 2176
506 rAlaGlyGlnGlySerLeuValThrValGlyLysGlyAlaHisSerIleS 523
2177 GTACACGTTCG.....GACTGGACGGGTCTGACAAGT 2208
523 erThrLeuGlyPheSerGlyGlyThrValAspPheGlyAlaLeuThrAla 539
2209 .....TGTACCGAAAAAACCATTACCGACGATAAAGTG..... 2241
540 GlyAlaGlnMetThrGluGlyThrValAsnValSerLysThrLeuAspLe 556
2242 .....ATTGCTTCATTGAGCAAGACCGACATCAGAGGCAATG 2278
556 uArgGlyGluGlyValIleGlnValSerAspSerAspValValArgSerV 573
2279 TCAGCCTTGCCGATCACGCTCATTTAAATCTCACA.....GGA 2316
573 alSerArgAspIleAspSerAlaLeuSerLeuThrGluValAspAspGly 589
2317 CTTGCCACACTCAAC..... 2331
590 AsnSerThrIleLysLeuValAspAlaGlnGlyAlaGluValLeuGlyAs 606
2332 ....GGCAATCTTAGTGCA.....GGCGGAGACACGCACTATA 2365
606 pAlaGlyAsnLeuGlnLeuGlnAspLysAsnGlyGlnIleLeuSerSerS 623
2366 CGGTTACGCGCAACGCCACCCAAAACGGCAACCTCAGCCTCGTGGGCAAT 2415
623 erAlaGlnArgAspIleGlnGlnAsnGlyGlnLysAlaAlaValGly... 638
2416 GCCCAAGCAACATTTAATCAAGCCACATTAAACGGCAACACATCGGCTTC 2465
639 .....ThrTyrAspTyrArgLeuThrSerGlyValAsnAsnAspGl 652
2466 GGACAATGCTTCATTTAATCTAAGCAAC.....AACGCCGTACAAA 2506
652 yLeuTyrIleGlyTyrGlyLeuThrGlnLeuAspLeuHisAlaThrAspS 669
2507 ACGGCAGTCTGACGCTTTCCGACAACGCTAAGGCAAACGTAAGCCATTCC 2556
669 erAspAlaLeuValLeuSerSerAsnGlyLys.....SerGluAsn 682
2557 GCACTCAACGGCAATGTCTCCCTAGCCGATAAGGCAGTATTCCATTTTGA 2606
683 AlaAlaAspLeuSerAlaLysIleThrGlySerGlyAspLeuAlaPheSe 699
2607 AAACAGCCGCTTTACCGGAAAAATCAGCGGCGGCAAGGATACGGCATTAC 2656
699 rSerGlnLys.....GlyGlnThrValSerLeuS 709
2657 ACTTAAAAGACAGCGAATGGACGCTGCCGTCGGGCACGGAATTAGGCAAT 2706
709 erAsnLysAspAsnAspTyrThrGlyValThrAspLeuArgSerGlyThr 725
2707 TTAAACCTTGACAACGCCACCATTACACTCAATTCCGCCTATCGACACGA 2756
726 LeuLeuLeuAsnAsnAspAsnValLeuGlyAsnThr.....HisGl 739
2757 TGCGGCAGGCGCGCAAACCGGCAGTGCGGCAGATGCGCCGCGCCGCCGTT 2806
739 u..... 739
2807 CGCGCCGTTCCCTATTATCCGTTACGCCGCCAACTTCGGCAGAATCCCGT 2856
740 .....LeuArgLeuAlaAlaGluThrGluLeuAspMetAsn 751


2857 TTCAACACGCTGACGGTAAACGGCAAATTGAACGGTCAGGGAACATTCCG 2906
752 GlyHisSerGlnThrVal...GlyThrLeuAsnGlySerAla..... 764
2907 CTTTATGTCGGAACTCTTCGGCTACCGCAGCGGCAAATTGAAGCTGGCGG 2956
765 .....AspSerLeuLeuSerLeuAsnGlyGlySerLeuThrValThrA 779
2957 AAAGTTCCGAAGGCACTTACACCTTGGCTGTCAACAATACCGGCAACGAA 3006
779 snGlyGlyThrSerThrGlySerLeu.....ThrGlySerGly 791
3007 CCCGTAAGTCTCGAGCAATTGACGGTAGTGGAAGGAAAAGACAACACACC 3056
792 GluLeuAsnIleGlnGlyGlyThrLeuAspIleAlaGlyAspAsnSerAs 808
3057 GCTGTCCGAAAATCTTAATTTC.....ACCCTGCAAAACG 3091
808 nLeuThrAlaAsnValAsnIleAlaAsnSerAlaAsnValLeuValSerH 825
3092 AACACGTCGATGCCGGCGCATGGCGTTATCAGCTTATCCGCAAAGACGGC 3141
825 isAlaGlnGlyLeuGlySerAlaAsn.....ValGluAsnAsnGly 838
3142 GAGTTCCGCCTGCATAATCCGGTCAAAGAACAA..... 3174
839 ThrLeuAlaLeuAsnAsnSerAlaGluLysArgAlaAlaAlaSerValAs 855
3175 ....GAGCTTTCCGACAAACTCGGCAAGGCGGGAGAAACAGAGGCCGCCT 3220
855 nTyrAlaLeuGlyGlyAsnLeuThrAsnAsnGlyThrLeuMetThrGlyM 872
3221 TGACGGCAAAACAGGCA.....CAACTTGCCGCCAAACAACAGGCGGAA 3264
872 etSerGlyGlnGlnAlaGlyAsnValLeuValValLysGlyAsnTyrHis 888
3265 AAAGACAACGCGCAAAGCCTT..... 3285
889 GlyAsnAsnGlyGlnLeuValMetAsnThrValLeuAsnGlyAspAspSe 905
3286 .....GACGCGCTGATTGCGGCCGGGCGCAATGCCACCGAAAAGGCAG 3328
905 rValThrAspLysLeuValValGluGly.....AspThrSerG 918
3329 AAAGTGTTGCCGAACCGGCCCGGCAGGCAGGCGGGGAAAATGCC..... 3372
918 lyThrThrAlaValThrValAsnAsnAlaGlyGlyThrGlyAlaLysThr 934
3373 .....GGCATT.....ATGCAGGCGGAGGAAGAGAAAAAACG 3404
935 LeuAsnGlyIleGluLeuIleHisValAspGlyLysSerGluGlyGluPh 951
3405 GGTGCAGGCGGATAAAGACACCGCCTT..... 3431
951 eValGlnAlaGlyArgIleValAlaGlyAlaTyrAspTyrThrLeuAlaA 968
3432 ..GGCGAAACAGCGCGAAGCGGAAAC.....CCGGCCGGC 3464
968 rgGlyGlnGlyAlaAsnSerGlyAsnTrpTyrLeuThrSerGlySerAsp 984
3465 TACCACCGCCTTCCCCCGCGCCCG.....CCGCGCCCGCC 3499
985 SerProGluLeuGlnProGluProAspProMetProAsnProGluPro.A 1001
3500 GGGATTTGCCGCAACCGCAGCCCCAACCGCAACCCCAACCGCAGCGCGAC 3549
1001 snProAsnProGluProAsnProAsnProThrProThrProGlyProAsp 1017
3550 CTG.....ATCAGCCGTTATGCCAATAGCGGTTTGAGTGA 3584
1018 LeuAsnValAspAsnAspLeuArgProGluAlaGlySerTyrIleAlaAs 1034
3585 ATTTTCCGCCACGCTCAACAGCGTTTTCGCC.....GTACAGGACGAAT 3628
1034 nLeuAlaAlaAla...AsnThrMetPheThrThrArgLeuHisGluArgL 1050
3629 TG.....GACCGCGTGTTTGCCGAAGACCGCCGCAAC 3660
1050 euGlyAsnThrTyrTyrThrAspMetValThrGlyGluGlnLysGlnThr 1066
3661 GCCGTTTGGACAAGCGGCATCCGGGACACCAAACACTACCGTTCGCAAGA 3710
1067 ThrMetTrpMetArg.....HisGluGlyGlyHisAsnLysTrpArgAs 1081
3711 TTTCCGCGCCTACCGCCAACAAACCGACCTGCGCCAAATCGGTATGCAGA 3760
1081 p..... 1081
3761 AAAACCTCGGCAGCGGGCGCGTCGGCATCCTGTTTTCGCACAACCGGACC 3810
1082 .....GlySerGlyGlnLeu.....LysThrGln 1089
3811 GGAAACACCTTCGACGACGGCATCGGCAACTCGGCACGGCTTGCCCACGG 3860
1090 SerAsnArgTyrValLeuGlnLeuGlyGlyAspVal..... 1101
3861 TGCCGTTTTCGGGCAATACGGCATCGGCAGGTTCGACATCGGCATCAGCG 3910
1102 .AlaGlnTrpSerGlnAsnGlySerAspArgTrpHisValGlyValMetA 1118
3911 CGGGCGCGGGTTTTAGTAGCGGC.....AGCCTTTCAGACGGCATC 3951
1118 laGlyTyrGlyAsnSerAspSerLysThrIleSerSerArgThrGlyTyr 1134
3952 AGAGGCAAAATCCGCCGCCGCGTGCTGCATTACGGCATTCAGGCAAGATA 4001
1135 ArgAlaLysAlaSerValAsnGlyTyrSerThrGlyLeuTyrAlaThrTr 1151
4002 CCGCGCAGGTTTCGGCGGATTCGGCATCGAACCGCACATCGGCGCAACGC 4051
1151 pTyrAla.....AspAspGluSerArgAsnGlyAla.... 1161
4052 GCTATTTCGTCCAAAAAGCGGATTACCGATACGAAAACGTCAATATCGCC 4101
1162 ..TyrLeuAspSerTrpAlaGlnTyrSerTrpPheAspAsnThrValLys 1177
4102 ACCCCGGGCCTTGCATTCAACCGCTACCGC.....GCGGG 4136
1178 GlyAspAspLeuGlnSerGluSerTyrLysSerLysGlyPheThrAlaSe 1194
4137 CATTAAGGCAGATTATTCATTCAAACCGGCGCAACACATTTCC..... 4179
1194 rLeuGluAlaGlyTyrLysHisLysLeuAlaGluPheAsnGlySerGlnG 1211
4180 .....ATCACGCCTTATTTGAGCCTGTCCTATACC 4209
1211 lyThrArgAsnGluTrpTyrValGlnProGlnAlaGlnValThrTrpMet 1227
4210 GATGCC..... 4215
1228 GlyValLysAlaAspLysHisArgGluSerAsnGlyThrLeuValHisSe 1244
4216 ....GCTTCCGGCAAAGTCCGAACGCGCGTCAATACCGCCGTATTGGCGC 4261
1244 rAsnGlyAspGlyAsnValGlnThrArgLeuGlyValLysThrTrpLeuL 1261
4262 AG.....GATTTCGGCAAAACCCGC..... 4281
1261 ysSerHisHisLysMetAspAspGlyLysSerArgGluPheGlnProPhe 1277
4282 ...AGTGCGGAATGGGGCGTAAACGCCGAAATCAAAGGTTTCACGCTGTC 4328
1278 ValGluValAsnTrpLeuHisAsnSer.....LysAspPheSerThrSe 1292
4329 CCTCCACGCTGCCGCCGCCAAGGGGCCGCAATTGGAAGCGCAGCACAGCG 4378

```
         |:::   :::::::::::         |||      |||:::::   |
1292 rMetAspGlyValSerVal......ThrGlnAspGlyAlaArgAsnIleA 1307

4379 CGGGCATCAAATTAGGC 4395
      ||   ||||||   ||||
1307 laGluIleLysThrGly 1312
```

seq_name: pir2:C91068

seq_documentation_block:
hypothetical protein ECs3515 [imported] - Escherichia coli (strain O157:H7, substrain RI
C;Species: Escherichia coli
C;Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 18-Jul-2001
C;Accession: C91068
R;Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.
gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.
DNA Res. 8, 11-22, 2001
A;Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and genc
A;Reference number: A99629; MUID:21156231; PMID:11258796
A;Accession: C91068
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1571 <HAY>
A;Cross-references: GB:BA000007; PIDN:BAB36938.1; PID:g13362986; GSPDB:GN00154
A;Experimental source: strain O157:H7, substrain RIMD 0509952
C;Genetics:
A;Gene: ECs3515

alignment_scores:

```
           Quality:  345.50          Length:   1543
             Ratio:  0.482             Gaps:     83
Percent Similarity:  46.468  Percent Identity: 20.609
```

alignment_block:
US-09-303-518D-653 x C91068 ..

Align seg 1/1 to: C91068 from: 1 to: 1571

```
 343 CATAACGGCGGCTATAACAATGTTGATTTGGTGCGGAGGGAAGCAATCC 392
      :::::||||||    :::|||:::  |||:::  ::::::|||
 308 TyrSerGlyGlyValLeuAspValSer...GlyGlyThrAlaThrAsnVa 323

 393 CGATCAGCACCGCTTTTCTTACCAAATTGTGAAAAGAAATAATTATAAAG 442
         ||||||       :::      |||:::|||  |||:::       :
 323 lThrGlnHis......AspGlyAlaIleLeuLysThrAsnThrAsnGlyT 338

 443 CAGGGACTAACGGCCATCCTTATGGCGGCGATTATCATATGCCGCGTTTG 492
      ::   ::::::|||           |||   :::           :::
 338 hrThrValSerGlyThrAsnSerGluGlyAlaPhe.........SerIle 351

 493 CACAAATTTGTCACAGATGCAGAACCTGTTGAGATGACCAGTTATATGGA 542
     |||:::   |||:::|||          :::|||      :::::::::||
 352 HisAsnHisValAlaAspAsnValLeuLeuGluAsnGlyGlyHisLeuAs 368

 543 TGGGTGGAAATACGCTGATTTAAATAAATACCCTGATC............ 580
     |           |||:::::::   |||||| |||:::|||
 368 pIleAsnAlaTyrGlySerAlaAsnLys.ThrIleIleLysAspLysGly 384

 581 .........GTGTTCGAATCGGAGCAGGCAGAC............AATAT 609
             |||:::   :::   :::||||||             :::
 385 ThrMetSerValLeuThrAsnAlaLysAlaAspAlaThrArgIleAspAs 401

 610 TGGCGGTCTGATGAAGACGAACCCAATAACCGCGAAAGTTCATATCATAT 659
      ||||||:::|||  |||         :::|||
 401 nGlyGlyValMet..AspValAlaGlyAsn.................... 410

 660 TGCAAGCGCATATTCTTGGCTCGTCGGTGGCAATACCTTTGCACAAAATG 709
            |||   :::   :::::::::|||:::|||          |||:
 411 .......AlaThrAsnThrIleIleAsnGlyGlyThrGlnAsnIleAsnA 425

 710 GATCAGGT.........GGTGGCACAGTCAACTTAGGTAGCGAAAAAATT 750
      ::   |||            |||   :::::::|||   |||:::::::::|||
 425 snTyrGlyIleAlaThrGlyThrAsnIleAsnSerGlyThrGlnAsnIle 441

 751 AAACATAGCCCATATGGTTTTTTACCAACAGGAGGC............TC 788
     |||                            :::||||||           ::
 442 Lys......................SerGlyGlyLysAlaAspThrTh 450

 789 ATTTGGCGACAGTGGCTCACCAATGTTTATCTATGATGCCCAAAAGCAAA 838
      ::::   :::|||||||||   :::   :::            :::|||:::
 450 rIleIleSerSerGlySerArgGlnValVal.........GluLysAspG 464

 839 AGTGGTTAATTAATGGGGTATTGCAAACAGGCAACCCCTATATAGGAAAA 888
              |||:::::   :::::::::|||:::        |||
 464 lyThrAlaIleGlySerAsnIleSerAlaGlyGlySerLeuIleValTyr 480

 889 AGCAATGGCTTCCAGCTAGTTCGTAAAGATTGGTTCTATGATGAAATCTT 938
      ::::::|||:::
 481 ThrGlyGlyIle...................................... 484

 939 TGCTGGAGATACCCATTCAGTATTCTACGAACCACATCAAAATGGGAAAT 988
               :::|||:::|||        |||         :::|||:::
 485 ..........AlaHisGlyValAsnGlnGlu.........ThrGlySerA 495

 989 ACTTTTTTAACGACAATAATAATGGCGCAGGA...AAAATCGATGCCAAA 1035
      :::                |||:::|||||||||    |||:::::
 495 laLeuVal......AlaAsnThrGlyAlaGlyThrAspIleGluGlyTyr 509

1036 CATAAA......CACTATTCTCTACCT..............TATAGATT 1064
     :::|||      |||:::::::::                |||   ::
 510 AsnLysLeuSerHisPheThrIleThrGlyGlyGluAlaAsnTyrValVa 526

1065 AAAAACACGAACCGTTCAATTGTTTAATGTTTCTTTATCCGAGACAGCAA 1114
      :      :::|||    :::|||        :::   :::::::::|||:::
 526 lLeuGluAsnThrGlyGluLeu......ThrValValAlaLysThrSerA 541

1115 GAGAACCTGTTTATCATGCTGCAGGTGGG..................... 1143
      :::    :::       |||||||||
 541 laLysAsnThrThrIleAspAlaGlyGlyLysLeuIleValGlnLysGlu 557

1144 GTCAACAGTTATCGACCCAGACTGAATAATGGAGAAAATATTTCCTTTAT 1193
     :::::::::          ||||||||||||||        ::::::
 558 AlaLysThrAspSerThrArgLeuAsnAsnGlyGlyValLeuGluValGl 574

1194 TGACAAAGGAAAAGGTGAATTGATACTTACCAGCAACATCAACCAAGGCG 1243
      |||   |||:::::::::              :::::::::|||   :
 574 nAspGlyGlyGluAlaLys..............HisValGluGlnGlnS 586

1244 CGGGCGGTTTGTATTTTGAGGGTAATTTTACGGTCTCGCCTAAAAACAAC 1293
      ::||||||      :::   ::::::   :::   :::      ::::::
 586 erGlyGlyAlaLeuIleAlaSerThrThrSerGlyThrLeuIleGluGly 602

1294 GAAACGTGGCAAGGCGCGGGCGTTCATATCAGTGATGGCAGTACCGTTAC 1343
        :::      |||   :::   :::|||   :::::::::
 603 ThrAsnSerTyrGlyAspAlaPheTyrIleArgAsnSerGlu........ 616

1344 TTGGAAAGTAAACGGCGTGGCAAACGACCGCCTGTCCAAAATCGGCAAAG 1393
                                               :::|||:
 617 .......................................AlaLysA 619

1394 GCACGCTGCTGGTTCAAGCCAAAGGGGAAAACCAAGGCTCGGTCAGCGTG 1443
      :::::::|||                 ||||||   ||||||::::::|||
 619 snValValLeu..............GluAsnAlaGlySerLeuThrVal 630

1444 GGCGACGGTAAAGTCATCTTAGATCAGCAGGCGGACGATCAAGGCAAAAA 1493
           |||:::       :::|||         :::   :::||||||
 631 ValThrGlySerArgAlaValAspThrIleIleAsnAlaAsnGlyLysMe 647

1494 ACAAGCCTTT...AGTGAAATCGGCTTGGTC.........AGCGGCAGGG 1531
      :::::::::   :::::::::|||   |||           :::|||
 647 tAspValTyrGlyLysAspValGlyThrValLeuAsnSerAlaGlyThrG 664

1532 GGACGGTGCAACTGAATGCCGATAATCAGTTCAACCCCGACAAACTCTAT 1581
      |||:::        :::|||   :::               ||||||
 664 lnThrIleTyrAlaSerAlaThrSer............AspLysAla... 675
```

```
1582 TTCGGCTTTCGCGGCGGACGTTTGGATTTGAACGGGCATTCGCTTTCGTT 1631
      ::::::::::||||||:::      :::   |||   :::   :::
 676 ...AsnIleLysGlyGlyLysGlnThrValTyrGlyLeuAlaThrGlu.. 690

1632 CCACCGCATTCAAAATACCGATGAAGGGGCGATGATTGTCAACCACAATC 1681
                |||   ::::::|||   :::||||||:::   ::::
 691 ..........AlaAsnIleGluSerGlyGluGlnIleValAspGlyGlyS 704

1682 AAGACAAAGAATCCACCGTTACCATTACAGGCAATAAAGATATTACTACA 1731
     ::   :::::::::   ::::::      |||::::::   :::   :::
 704 erThrGluLysThrHisIleAsn...GlyGlyThrGlnThrValGlnAsn 719

1732 ACCGGCAATAACAACAACTTGGAT.........AGCAAAAAAGAAATTGC 1772
        |||:::      |||   |||          :::   :::::|||
 720 TyrGlyLysAlaIleAsnThrAspIleValSerGlyLeuGlnGlnIleMe 736

1773 CTACAACGGTTGGTTTGGCGAGAAAGATGCAACCAAAACGAACGGGCGGC 1822
        |||          |||       :::::::::       ||||||
 736 tAlaAsn.........GlyThrAlaGluGlySerIleIleAsnGlyGlyS 750

1823 TCAATCTGAATTACCAACCGGAAGAAGCGGATCGCACTTTACTGCTTTCC 1872
       ::::::       :::          |||:::::::::::|||   :::
 750 erGlnIleValAsnGluGlyGlyLeuAlaGluAsnSerValLeuAsnAsp 766

1873 GGCGGAACA.....................AATTTAAACGGCAATATCAC 1901
     |||||||||                    :::   :::|||
 767 GlyGlyThrLeuAspValArgGluLysGlySerAlaThrGlyIleGlnGl 783

1902 GCAAACAAACGGCAAACTGTTTTTCAGCGGCAGACCGACACCGCACGCCT 1951
      ::::::::::|||   |||       :::   |||   |||       :::
 783 nSerSerGlnGlyAlaLeuValAlaThrThrArgAlaThrArgValThr. 799

1952 ACAATCATTTAGGAAGCGGGTGGTCAAAAATGGAAGGTATCCCACAAGGA 2001
                              ::::::   :::|||:::
 800 ...................GlyThrArgAlaAspGlyValAla...... 807

2002 GAAATCGTGTGGGACAACGATTGGATCGACCGCACATTTAAAGCGGAAAA 2051

 807 ..................................................... 807

2052 CTTCCATATTCAGGGCGGACAAGCG.............GTGGTTTCCCGCA 2089
      |||   |||:::   |||   |||             :::::::::   :
 808 .PheSerIleGluGlnGlyAlaAlaAsnAsnIleLeuLeuAlaAsnGlyG 824

2090 ATGTTGCCAAAGTGGAAGGCGATTGGCATTTAAGCAATCACGCCCAAGCA 2139
     ::|||      ||||||:::|||
 824 lyValLeuThrValGluSerAsp........................... 831

2140 GTTTTCGGTGTCGCACCGCATCAAAGCCACACAATCTGTACACGTTCGGA 2189
                                        |||   |||||
 832 .....................................ThrSerSerAs 835

2190 CTGGACGGGTCTGACAAGTTGTACCGAAAAAACCATTACCGACGATAAAG 2239
     |   |||   :::::::::       ::::::   ::          |||:
 835 pLysThrGlnValAsnThrGlyGlyArgGluIleVal.........LysT 849

2240 TGATTGCTTCATTGAGCAAGACCGACATCAGAGGCAATGTCAGCCTT... 2286
     ::   |||:::   :::   |||   :::   |||:::   ::::::
 849 hrLysAlaThrAlaThrGlyThrThrLeuThrGlyGlyGluGlnIleVal 865

2287 .........GCCGATCACGCTCATTTAAATCTCACAGGACTTGCCACACT 2327
             |||::::::::::   :::|||      |||:::   |||::
 866 GluGlyValAlaAsnGluThrThrIleAsnAspGlyGlyIleGlnThrVa 882

2328 CAACGGCAAT.....................CTTAGTGCAGGCGGAGACA 2356
     ::::::::|||                    ::::::   ||||||
 882 lSerAlaAsnGlyGluAlaIleLysThrThrIleAsnGluGlyGly.... 897

2357 CGCACTATACGGTTACGCGCAACGCCACCCAAAACGGCAACCTCAGC... 2403
            |||:::|||   |||       :::|||||:::   :::
 898 ........ThrLeuThrValAsn......AspAsnGlyLysAlaThrAsp 909

2404 CTCGTGGGCAATGCCCAAGCAACATTTAATCAAGCCACATTAAACGGC.. 2451
     :::|||   |||:::   |||:::::::::   :::|||   ||||||
  910 IleValGlnAsnSerGlyAlaAlaLeuGlnThrSerThrAlaAsnGlyIl 926

2452 ....AACACATCGGCTTCGGACAATGCTTCATTTAATCTAAGCAACAACG 2497
         :::::::::   :::   :::::::|||:::::::|||:::|||
  926 eGluIleSerGlyThrHisGlnTyrGlyThrPheSerIleSerGlyAsnL 943

2498 CCGTACAAAAC.................GGCAGTCTGACGCTTTCCGAC 2529
       :::   |||                 |||:::|||   :::
  943 euAlaThrAsnMetLeuLeuGluAsnGlyGlyAsnLeuLeuValLeuAla 959

2530 AACGCTAAGGCAAACGTAAGCCATTCCGCACTCAACGGCAATGTCTCCCT 2579
     :::::::::|||:::   |||      :::   :::|||
  960 GlyThrGluAlaArgAspSerThrValGlyLysGlyGly........... 972

2580 AGCCGATAAGGCAGTATTCCATTTTGAAAACAGCCGCTTTACCGGAAAAA 2629
               |||:::   :::      ::::::      |||
  973 ..........AlaMetGlnAsnGlnGlyGlnAspSerAlaThrLysValA 986

2630 TCAGCGGCGGCAAGGATACGGCATTACACTTAAAAGACAGCGAATGGACG 2679
       |||||||||:::  |||      :::   ||||||:::       :::
  986 snSerGlyGlyGlnTyrThrLeuGlyArgSerLysAspGluPheGlnAla 1002

2680 CTGCCGTCGGGCACGGAATTA.....................:..... 2700
     |||       :::   :::|||
 1003 LeuAlaArgAlaGluAspLeuGlnValAlaGlyGlyThrAlaIleValTy 1019

2701 .....................................GGCAATTTAA 2710
                                         |||:::|||:
 1019 rAlaGlyThrLeuAlaAspAlaSerValSerGlyAlaThrGlySerLeuS 1036

2711 ACCTT............GACAACGCCACC...ATTACACTCAATTCCGCC 2745
      ::|||           ||||||:::|||   :::   |||::::::|||
 1036 erLeuMetThrProArgAspAsnValThrProValLysLeuGluGlyAla 1052

2746 TATCGA...CACGATGCGGCAGGCGCGCAAACCGGCAGTGCGGCAGAT.. 2790
        |||      |||:::|||           |||::::::::::|||
 1053 IleArgIleThrAspSerAlaThrLeuThrIleGlyAsnGlyValAspTh 1069

2791 ..................GCGCCGCGCCGCCGTTCG............. 2808
                         |||      |||   |||
 1069 rThrLeuAlaAspLeuThrAlaAlaSerArgGlySerValTrpLeuAsnS 1086

2809 .....................................CGCCGTTCC 2817
                                          :::|||
 1086 erAsnAsnSerCysAlaGlyThrSerAsnCysGluTyrArgValAsnSer 1102

2818 CTATTA............................TCCGTTACGCCGCC 2837
     ||||||                          :::::::::   ||
 1103 LeuLeuLeuAsnAspGlyAsnValTyrLeuSerAlaGlnThrAlaAlaPr 1119

2838 AACTTCGGCAGAATCCCGTTTCAACACGCTGACGGTAAACGGCAAATTGA 2887
     |:::::::::::::::   :::||||||||||||:::|||   :::|||:
 1119 oAlaThrThrAsnGlyIleTyrAsnThrLeuThrThrAsn...GluLeuS 1135

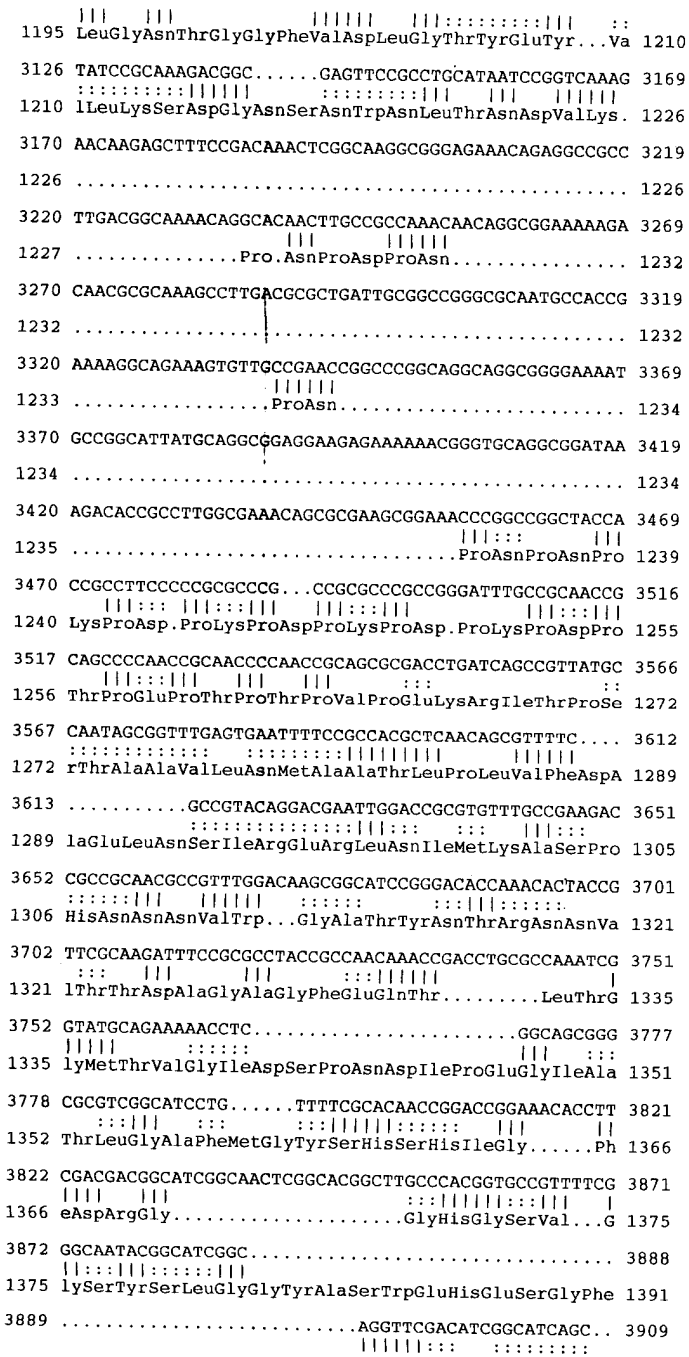
2888 ACGGTCAGGGAACATTCCGCTTTATGTCGGAACTCTTCGGCTACCGCAGC 2937
     ::|||:::|||:::|||   :::   :::::::::   |||   |||:::
 1135 erGlySerGlyAsnPheTyrLeuHisThrAsnValAlaGlySerArgGly 1151

2938 GGCAAATTGAAGCTGGCGGAAAGTTCCGAAGGCACTTACACCTTGGCTGT 2987
        :::|||   :::   :::::::::   |||::::::   :::   ||
 1152 AspGlnLeuValValAsnAsnAsnAlaThrGlyAsnPheLysIlePheVa 1168

2988 CAACAATACCGGCAACGAACCCGTAAGTCTCGAGCAATTGACGGTAGTGG 3037
     |:::::|||||   :::|||   |||   :::   :::|||:::||||:
 1168 lGlnAspThrGlyValSerProGlnSerAspAspAlaMetThrLeuValL 1185

3038 AA......GGAAAAGACAACACACCGCTGTCCGAAAATCTTAATTTCACC 3081
     ::      |||   |||                          :::|||:::
 1185 ysThrGlyGlyGlyAsp.....................AlaSerPheSer 1194

3082 CTGCAAAACGAA......CACGTCGATGCCGGCGCATGGCGTTATCAGCT 3125
```



```
1195 LeuGlyAsnThrGlyGlyPheValAspLeuGlyThrTyrGluTyr...Va 1210

3126 TATCCGCAAAGACGGC......GAGTTCCGCCTGCATAATCCGGTCAAAG 3169
1210 lLeuLysSerAspGlyAsnSerAsnTrpAsnLeuThrAsnAspValLys. 1226

3170 AACAAGAGCTTTCCGACAAACTCGGCAAGGCGGGAGAAACAGAGGCCGCC 3219

1226 ..................................................... 1226

3220 TTGACGGCAAAACAGGCACAACTTGCCGCCAAACAACAGGCGGAAAAAGA 3269
1227 ..............Pro.AsnProAspProAsn..................... 1232

3270 CAACGCGCAAAGCCTTGACGCGCTGATTGCGGCCGGGCGCAATGCCACCG 3319

1232 ..................................................... 1232

3320 AAAAGGCAGAAAGTGTTGCCGAACCGGCCCGGCAGGCAGGCGGGGAAAAT 3369
1233 ..................ProAsn............................. 1234

3370 GCCGGCATTATGCAGGCGGAGGAAGAGAAAAAACGGGTGCAGGCGGATAA 3419

1234 ..................................................... 1234

3420 AGACACCGCCTTGGCGAAACAGCGCGAAGCGGAAACCCGGCCGGCTACCA 3469
1235 ....................................ProAsnProAsnPro 1239

3470 CCGCCTTCCCCCGCGCCCG...CCGCGCCCGCCGGGATTTGCCGCAACCG 3516
1240 LysProAsp.ProLysProAspProLysProAsp.ProLysProAspPro 1255

3517 CAGCCCCAACCGCAACCCCAACCGCAGCGCGACCTGATCAGCCGTTATGC 3566
1256 ThrProGluProThrProThrProValProGluLysArgIleThrProSe 1272

3567 CAATAGCGGTTTGAGTGAATTTTCCGCCACGCTCAACAGCGTTTTC.... 3612
1272 rThrAlaAlaValLeuAsnMetAlaAlaThrLeuProLeuValPheAspA 1289

3613 ..........GCCGTACAGGACGAATTGGACCGCGTGTTTGCCGAAGAC 3651
1289 laGluLeuAsnSerIleArgGluArgLeuAsnIleMetLysAlaSerPro 1305

3652 CGCCGCAACGCCGTTTGGACAAGCGGCATCCGGGACACCAAACACTACCG 3701
1306 HisAsnAsnAsnValTrp...GlyAlaThrTyrAsnThrArgAsnAsnVa 1321

3702 TTCGCAAGATTTCCGCGCCTACCGCCAACAAACCGACCTGCGCCAAATCG 3751
1321 lThrThrAspAlaGlyAlaGlyPheGluGlnThr.........LeuThrG 1335

3752 GTATGCAGAAAAACCTC.....................GGCAGCGGG 3777
1335 lyMetThrValGlyIleAspSerProAsnAspIleProGluGlyIleAla 1351

3778 CGCGTCGGCATCCTG......TTTTCGCACAACCGGACCGGAAACACCTT 3821
1352 ThrLeuGlyAlaPheMetGlyTyrSerHisSerHisIleGly......Ph 1366

3822 CGACGACGGCATCGGCAACTCGGCACGGCTTGCCCACGGTGCCGTTTTCG 3871
1366 eAspArgGly.....................GlyHisGlySerVal...G 1375

3872 GGCAATACGGCATCGGC................................ 3888
1375 lySerTyrSerLeuGlyGlyTyrAlaSerTrpGluHisGluSerGlyPhe 1391

3889 .........................AGGTTCGACATCGGCATCAGC.. 3909

1392 TyrLeuAspGlyValValLysLeuArgPheGluSerAsnValAlaGl 1408

3910 .......GCGGGCGCGGGTTTTAGTAGCGGCAGCCTT...TCAGACGGCA 3949
1408 yLysMetSerSerGlyGlyAlaAlaAsnGlySerTyrHisSerAsnGlyL 1425

3950 TCAGAGGCAAAATCCGCCGCCGCGTGCTGCATTACGGCATTCAGGCAAGA 3999
1425 euGlyGlyHis.........................IleGluThrGly 1432

4000 TACCGCGCAGGTTTCGGCGGATTCGGCATCGAACCGCACATCGGCGCAAC 4049
1433 MetArgPheThrAspGlyAsnTrpAsnLeuThrProTyrAlaSerLeuTh 1449

4050 GCGCTATTTCGTCCAAAAAGCGGATTACCGA............TACGAAA 4087
1449 rGlyPheThrAlaAspAsnProGluTyrHisLeuSerAsnGlyMetGluS 1466

4088 ACGTCAATATCGCCACCCCGGGCCTTGCATTCAACCGCTACCGC...GCG 4134
1466 erLysSerValAspThrArgSerIle............TyrArgGluLeu 1478

4135 GGCATTAAGGCAGATTATTCATTCAAACCGGCGCAACACATTTCCATCAC 4184
1479 GlyAlaThrLeuSerTyrAsnMetArgLeuGlyAsnGlyMetGluIleGl 1495

4185 GCCTTATTTGAGCCTGTCCTATACCGATGCCGCTTCCGGCAAAGTCCGAA 4234
1495 uProTrpLeuLysAlaAlaValArgLysGluPheValAspAspAsnArgV 1512

4235 CGCGCGTCAATACC..................................... 4248
1512 alLysValAsnAsnAspGlyAsnPheValAsnAspLeuSerGlyArgArg 1528

4249 ............GCCGTATTGGCGCAGGATTTCGGCAAAACCCGCAGTGC 4286
1529 GlyIleTyrGlnAlaGlyIleLysAlaSerPheSerSerThrLeuSerGl 1545

4287 GGAATGGGGCGTAAACGCCGAAATCAAAGGTTTCACGCTGTCCCTCCACG 4336
1545 yHisLeuGlyVal...............GlyTyrSer.........HisG 1554

4337 CTGCCGCCGCCAAGGGGCCG 4356
1554 lyAlaGlyValGluSerPro 1560
```

seq_name: pir2:D85912

seq_documentation_block:
hypothetical protein ypjA [imported] - Escherichia coli (strain O157:H7, substrain ED
C;Species: Escherichia coli
C;Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 14-Sep-2001
C;Accession: D85912
R;Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; May
iller, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apoda
Nature 409, 529-533, 2001
A;Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
A;Reference number: A85480; MUID:21074935; PMID:11206551
A;Accession: D85912
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1528 <STO>
A;Cross-references: GB:AE005174; NID:g12517083; PIDN:AAG57760.1; GSPDB:GN00145; UWGP:
A;Experimental source: strain O157:H7, substrain EDL933
C;Genetics:
A;Gene: ypjA

alignment_scores:

Quality:	344.50	Length:	1543
Ratio:	0.480	Gaps:	84
Percent Similarity:	46.533	Percent Identity:	20.609

```
alignment_block:
 US-09-303-518D-653 x D85912  ..

 Align seg 1/1  to: D85912  from: 1  to: 1528

     343 CATAACGGCGGCTATAACAATGTTGATTTTGGTGCGGAGGGAAGCAATCC 392
         ::::::||||||      :::|||:::   |||:::   :::::::|||
     265 TyrSerGlyGlyValLeuAspValSer...GlyGlyThrAlaThrAsnVa 280

     393 CGATCAGCACCGCTTTTCTTACCAAATTGTGAAAAGAAATAATTATAAAG 442
             ||||||      :::      |||:::|||   |||:::      :
     280 lThrGlnHis......AspGlyAlaIleLeuLysThrAsnThrAsnGlyT 295

     443 CAGGGACTAACGGCCATCCTTATGGCGGCGATTATCATATGCCGCGTTTG 492
         ::   ::::::|||            |||  :::            :::
     295 hrThrValSerGlyThrAsnSerGluGlyAlaPhe.........SerIle 308

     493 CACAAATTTGTCACAGATGCAGAACCTGTTGAGATGACCAGTTATATGGA 542
         |||:::   |||:::|||          :::|||      ::::::::::||
     309 HisAsnHisValAlaAspAsnValLeuLeuGluAsnGlyGlyHisLeuAs 325

     543 TGGGTGGAAATACGCTGATTTAAATAAATACCCTGATC............ 580
         |         |||::::::   |||||| |||:::|||
     325 pIleAsnAlaTyrGlySerAlaAsnLys.ThrIleIleLysAspLysGly 341

     581 .........GTGTTCGAATCGGAGCAGGCAGAC............AATAT 609
                 |||:::   :::   :::||||||           :::
     342 ThrMetSerValLeuThrAsnAlaLysAlaAspAlaThrArgIleAspAs 358

     610 TGGCGGTCTGATGAAGACGAACCCAATAACCGCGAAAGTTCATATCATAT 659
         ||||||:::|||  |||      :::|||
     358 nGlyGlyValMet..AspValAlaGlyAsn.................... 367

     660 TGCAAGCGCATATTCTTGGCTCGTCGGTGGCAATACCTTTGCACAAAATG 709
                 |||   :::   :::::::::|||:::|||          |||:
     368 .......AlaThrAsnThrIleIleAsnGlyGlyThrGlnAsnIleAsnA 382

     710 GATCAGGT.........GGTGGCACAGTCAACTTAGGTAGCGAAAAAATT 750
         ::   |||          |||   :::::::|||   |||:::::::::|||
     382 snTyrGlyIleAlaThrGlyThrAsnIleAsnSerGlyThrGlnAsnIle 398

     751 AAACATAGCCCATATGGTTTTTTACCAACAGGAGGC............TC 788
         |||                          :::|||||||           ::
     399 Lys.......................SerGlyGlyLysAlaAspThrTh 407

     789 ATTTGGCGACAGTGGCTCACCAATGTTTATCTATGATGCCCAAAAGCAAA 838
         ::::   :::|||||||||   :::   :::          :::|||:::
     407 rIleIleSerSerGlySerArgGlnValVal.........GluLysAspG 421

     839 AGTGGTTAATTAATGGGGTATTGCAAACAGGCAACCCCTATATAGGAAAA 888
               |||::::::   :::::::::|||:::      |||
     421 lyThrAlaIleGlySerAsnIleSerAlaGlyGlySerLeuIleValTyr 437

     889 AGCAATGGCTTCCAGCTAGTTCGTAAAGATTGGTTCTATGATGAAATCTT 938
         :::::::|||:::
     438 ThrGlyGlyIle..................................... 441

     939 TGCTGGAGATACCCATTCAGTATTCTACGAACCACATCAAAATGGGAAAT 988
                 :::|||:::|||      |||          :::|||:::
     442 .........AlaHisGlyValAsnGlnGlu.........ThrGlySerA 452

     989 ACTTTTTTAACGACAATAATAATGGCGCAGGA...AAAATCGATGCCAAA 1035
         :::              |||:::|||||||||      |||::::::
     452 laLeuVal......AlaAsnThrGlyAlaGlyThrAspIleGluGlyTyr 466

    1036 CATAAA......CACTATTCTCTACCT..............TATAGATT 1064
         :::|||      |||:::::::::                  |||   ::
     467 AsnLysLeuSerHisPheThrIleThrGlyGlyGluAlaAsnTyrValVa 483

    1065 AAAAACACGAACCGTTCAATTGTTTAATGTTTCTTTATCCGAGACAGCAA 1114
         :      :::|||   :::|||      :::   ::::::::::|||:::
     483 lLeuGluAsnThrGlyGluLeu......ThrValValAlaLysThrSerA 498

    1115 GAGAACCTGTTTATCATGCTGCAGGTGGG..................... 1143
           :::   :::          |||||||||
     498 laLysAsnThrThrIleAspAlaGlyGlyLysLeuIleValGlnLysGlu 514

    1144 GTCAACAGTTATCGACCCAGACTGAATAATGGAGAAAATATTTCCTTTAT 1193
         :::::::::         ||||||||||||||||      ::::::
     515 AlaLysThrAspSerThrArgLeuAsnAsnGlyGlyValLeuGluValGl 531

    1194 TGACAAAGGAAAAGGTGAATTGATACTTACCAGCAACATCAACCAAGGCG 1243
          |||   |||::::::::::                 ::::::::::|||   :
     531 nAspGlyGlyGluAlaLys..............HisValGluGlnGlnS 543

    1244 CGGGCGGTTTGTATTTTGAGGGTAATTTTACGGTCTCGCCTAAAAACAAC 1293
         ::||||||      :::   ::::::   :::   :::      ::::::
     543 erGlyGlyAlaLeuIleAlaSerThrThrSerGlyThrLeuIleGluGly 559

    1294 GAAACGTGGCAAGGCGCGGGCGTTCATATCAGTGATGGCAGTACCGTTAC 1343
            :::      |||   :::   :::|||   :::::::::
     560 ThrAsnSerTyrGlyAspAlaPheTyrIleArgAsnSerGlu........ 573

    1344 TTGGAAAGTAAACGGCGTGGCAAACGACCGCCTGTCCAAAATCGGCAAAG 1393
                                                 :::|||:
     574 ........................................AlaLysA 576

    1394 GCACGCTGCTGGTTCAAGCCAAAGGGGAAAACCAAGGCTCGGTCAGCGTG 1443
         :::::::::|||                ||||||   ||||||::::::|||
     576 snValValLeu..............GluAsnAlaGlySerLeuThrVal 587

    1444 GGCGACGGTAAAGTCATCTTAGATCAGCAGGCGGACGATCAAGGCAAAAA 1493
               |||:::      :::|||          :::   :::||||||
     588 ValThrGlySerArgAlaValAspThrIleIleAsnAlaAsnGlyLysMe 604

    1494 ACAAGCCTTT...AGTGAAATCGGCTTGGTC.........AGCGGCAGGG 1531
          :::::::::   :::::::::|||   |||          :::|||
     604 tAspValTyrGlyLysAspValGlyThrValLeuAsnSerAlaGlyThrG 621

    1532 GGACGGTGCAACTGAATGCCGATAATCAGTTCAACCCCGACAAACTCTAT 1581
           |||:::      :::|||   :::             ||||||
     621 lnThrIleTyrAlaSerAlaThrSer............AspLysAla... 632

    1582 TTCGGCTTTCGCGGCGGACGTTTGGATTTGAACGGGCATTCGCTTTCGTT 1631
            :::::::::||||||:::      :::   |||   :::   :::
     633 ...AsnIleLysGlyGlyLysGlnThrValTyrGlyLeuAlaThrGlu.. 647

    1632 CCACCGCATTCAAAATACCGATGAAGGGGCGATGATTGTCAACCACAATC 1681
                        |||   :::::::|||   :::|||||||:::   ::::
     648 ..........AlaAsnIleGluSerGlyGluGlnIleValAspGlyGlyS 661

    1682 AAGACAAAGAATCCACCGTTACCATTACAGGCAATAAAGATATTACTACA 1731
         ::   :::::::::   ::::::      |||::::::   :::   :::
     661 erThrGluLysThrHisIleAsn...GlyGlyThrGlnThrValGlnAsn 676

    1732 ACCGGCAATAACAACAACTTGGAT.........AGCAAAAAAGAAATTGC 1772
            |||:::      |||   |||          :::   ::::::|||
     677 TyrGlyLysAlaIleAsnThrAspIleValSerGlyLeuGlnGlnIleMe 693

    1773 CTACAACGGTTGGTTTGGCGAGAAAGATGCAACCAAAACGAACGGGCGGC 1822
          |||          |||      :::::::::      ||||||
     693 tAlaAsn.........GlyThrAlaGluGlySerIleIleAsnGlyGlyS 707

    1823 TCAATCTGAATTACCAACCGGAAGAAGCGGATCGCACTTTACTGCTTTCC 1872
           ::::::      :::          |||:::::::::::::|||   :::
     707 erGlnIleValAsnGluGlyGlyLeuAlaGluAsnSerValLeuAsnAsp 723

    1873 GGCGGAACA.....................AATTTAAACGGCAATATCAC 1901
         |||||||||                        :::   :::|||
     724 GlyGlyThrLeuAspValArgGluLysGlySerAlaThrGlyIleGlnGl 740

    1902 GCAAACAAACGGCAAACTGTTTTTCAGCGGCAGACCGACACCGCACGCCT 1951
          :::::::::|||   |||      :::   |||   |||      :::
     740 nSerSerGlnGlyAlaLeuValAlaThrThrArgAlaThrArgValThr. 756

    1952 ACAATCATTTAGGAAGCGGGTGGTCAAAAATGGAAGGTATCCCACAAGGA 2001
```

```
                                  ::::::   :::|||:::
 757 ....................GlyThrArgAlaAspGlyValAla..... 764

2002 GAAATCGTGTGGGACAACGATTGGATCGACCGCACATTTAAAGCGGAAAA 2051

 764 ..................................................... 764

2052 CTTCCATATTCAGGGCGGACAAGCG............GTGGTTTCCCGCA 2089
      |||   |||:::   |||  |||             :::::::::   :
 765 .PheSerIleGluGlnGlyAlaAlaAsnAsnIleLeuLeuAlaAsnGlyG 781

2090 ATGTTGCCAAAGTGGAAGGCGATTGGCATTTAAGCAATCACGCCCAAGCA 2139
     ::|||      ||||||:::|||
 781 lyValLeuThrValGluSerAsp.......................... 788

2140 GTTTTCGGTGTCGCACCGCATCAAAGCCACACAATCTGTACACGTTCGGA 2189
                                                |||   |||||
 789 .....................................ThrSerSerAs 792

2190 CTGGACGGGTCTGACAAGTTGTACCGAAAAAACCATTACCGACGATAAAG 2239
     |   |||   :::::::::      ::::::   :::         |||:
 792 pLysThrGlnValAsnThrGlyGlyArgGluIleVal.........LysT 806

2240 TGATTGCTTCATTGAGCAAGACCGACATCAGAGGCAATGTCAGCCTT... 2286
     ::   |||:::   :::   |||   :::   |||:::   ::::::
 806 hrLysAlaThrAlaThrGlyThrThrLeuThrGlyGlyGluGlnIleVal 822

2287 .........GCCGATCACGCTCATTTAAATCTCACAGGACTTGCCACACT 2327
             |||::::::::::   :::|||      |||:::   |||::
 823 GluGlyValAlaAsnGluThrThrIleAsnAspGlyGlyIleGlnThrVa 839

2328 CAACGGCAAT.....................CTTAGTGCAGGCGGAGACA 2356
     :::::::|||                     ::::::   ||||||
 839 lSerAlaAsnGlyGluAlaIleLysThrThrIleAsnGluGlyGly.... 854

2357 CGCACTATACGGTTACGCGCAACGCCACCCAAAACGGCAACCTCAGC... 2403
           |||:::|||   |||      :::|||||:::   :::
 855 ........ThrLeuThrValAsn......AspAsnGlyLysAlaThrAsp 866

2404 CTCGTGGGCAATGCCCAAGCAACATTTAATCAAGCCACATTAAACGGC.. 2451
     :::|||   |||:::   |||:::::::::   :::|||   ||||||
 867 IleValGlnAsnSerGlyAlaAlaLeuGlnThrSerThrAlaAsnGlyIl 883

2452 ....AACACATCGGCTTCGGACAATGCTTCATTTAATCTAAGCAACAACG 2497
          :::::::::   :::   ::::::|||:::::::|||:::|||
 883 eGluIleSerGlyThrHisGlnTyrGlyThrPheSerIleSerGlyAsnL 900

2498 CCGTACAAAAC.................GGCAGTCTGACGCTTTCCGAC 2529
       :::   |||                 |||:::|||   :::
 900 euAlaThrAsnMetLeuLeuGluAsnGlyGlyAsnLeuLeuValLeuAla 916

2530 AACGCTAAGGCAAACGTAAGCCATTCCGCACTCAACGGCAATGTCTCCCT 2579
     ::::::::::|||:::   |||      :::   :::|||
 917 GlyThrGluAlaArgAspSerThrValGlyLysGlyGly........... 929

2580 AGCCGATAAGGCAGTATTCCATTTTGAAAACAGCCGCTTTACCGGAAAAA 2629
           |||:::   :::      ::::::      |||
 930 .........AlaMetGlnAsnGlnGlyGlnAspSerAlaThrLysValA 943

2630 TCAGCGGCGGCAAGGATACGGCATTACACTTAAAAGAC............ 2667
      |||||||||:::   |||      :::   ||||||
 943 snSerGlyGlyGlnTyrThrLeuGlyArgSerLysAspGluPheGlnAla 959

2668 ............AGCGAATGGACGCTGCCGTCGGGCACGGAATTA..... 2700
                ::::::      :::   :::|||||||   :::
 960 ***AlaArgAlaGluAspLeuGlnValAlaGlyGlyThrAlaIleValTy 976

2701 .....................................GGCAATTTAA 2710
                                         |||:::|||:
 976 rAlaGlyThrLeuAlaAspAlaSerValSerGlyAlaThrGlySerLeuS 993

2711 ACCTT............GACAACGCCACC...ATTACACTCAATTCCGCC 2745
     ::|||           ||||||:::|||   :::   |||:::::::|||

 993 erLeuMetThrProArgAspAsnValThrProValLysLeuGluGlyAla 1009

2746 TATCGA...CACGATGCGGCAGGCGCGCAAACCGGCAGTGCGGCAGAT.. 2790
        |||      |||:::|||            |||:::::::::|||
1010 IleArgIleThrAspSerAlaThrLeuThrIleGlyAsnGlyValAspTh 1026

2791 ..................GCGCCGCGCCGCCGTTCG............. 2808
                         |||      |||  |||
1026 rThrLeuAlaAspLeuThrAlaAlaSerArgGlySerValTrpLeuAsnS 1043

2809 .....................................CGCCGTTCC 2817
                                          :::|||
1043 erAsnAsnSerCysAlaGlyThrSerAsnCysGluTyrArgValAsnSer 1059

2818 CTATTA..........................TCCGTTACGCCGCC 2837
     ||||||                        :::::::::   ||
1060 LeuLeuLeuAsnAspGlyAsnValTyrLeuSerAlaGlnThrAlaAlaPr 1076

2838 AACTTCGGCAGAATCCCGTTTCAACACGCTGACGGTAAACGGCAAATTGA 2887
     |:::::::::::::::   :::|||||||||||||:::|||   :::|||:
1076 oAlaThrThrAsnGlyIleTyrAsnThrLeuThrThrAsn...GluLeuS 1092

2888 ACGGTCAGGGAACATTCCGCTTTATGTCGGAACTCTTCGGCTACCGCAGC 2937
     ::|||:::|||:::|||   :::   :::::::::   |||   |||:::
1092 erGlySerGlyAsnPheTyrLeuHisThrAsnValAlaGlySerArgGly 1108

2938 GGCAAATTGAAGCTGGCGGAAAGTTCCGAAGGCACTTACACCTTGGCTGT 2987
        :::|||   :::   :::::::::   |||::::::   :::   ||
1109 AspGlnLeuValValAsnAsnAsnAlaThrGlyAsnPheLysIlePheVa 1125

2988 CAACAATACCGGCAACGAACCCGTAAGTCTCGAGCAATTGACGGTAGTGG 3037
     |::::::||||||   :::|||   |||   :::   :::|||:::|||:
1125 lGlnAspThrGlyValSerProGlnSerAspAspAlaMetThrLeuValL 1142

3038 AA......GGAAAAGACAACACACCGCTGTCCGAAAATCTTAATTTCACC 3081
     ::      |||   |||                          :::|||:::
1142 ysThrGlyGlyGlyAsp.....................AlaSerPheSer 1151

3082 CTGCAAAACGAA......CACGTCGATGCCGGCGCATGGCGTTATCAGCT 3125
     |||   |||              ||||||   |||:::::::::|||   ::
1152 LeuGlyAsnThrGlyGlyPheValAspLeuGlyThrTyrGluTyr...Va 1167

3126 TATCCGCAAAGACGGC......GAGTTCCGCCTGCATAATCCGGTCAAAG 3169
     ::::::::::||||||      ::::::::::|||   |||   ||||||
1167 lLeuLysSerAspGlyAsnSerAsnTrpAsnLeuThrAsnAspValLys. 1183

3170 AACAAGAGCTTTCCGACAAACTCGGCAAGGCGGGAGAAACAGAGGCCGCC 3219

1183 ..................................................... 1183

3220 TTGACGGCAAAACAGGCACAACTTGCCGCCAAACAACAGGCGGAAAAAGA 3269
                          |||      ||||||
1184 ..............Pro.AsnProAspProAsn..................... 1189

3270 CAACGCGCAAAGCCTTGACGCGCTGATTGCGGCCGGGCGCAATGCCACCG 3319

1189 ..................................................... 1189

3320 AAAAGGCAGAAAGTGTTGCCGAACCGGCCCGGCAGGCAGGCGGGGAAAAT 3369
                       ||||||
1190 ..................ProAsn............................ 1191

3370 GCCGGCATTATGCAGGCGGAGGAAGAGAAAAAACGGGTGCAGGCGGATAA 3419

1191 ..................................................... 1191

3420 AGACACCGCCTTGGCGAAACAGCGCGAAGCGGAAACCCGGCCGGCTACCA 3469
                                      |||:::   |||
1192 ................................ProAsnProAsnPro 1196

3470 CCGCCTTCCCCCGCGCCCG...CCGCGCCCGCCGGGATTTGCCGCAACCG 3516
        |||::: |||:::|||   |||:::|||          |||:::|||
1197 LysProAsp.ProLysProAspProLysProAsp.ProLysProAspPro 1212
```

```
3517 CAGCCCCAACCGCAACCCCAACCGCAGCGCGACCTGATCAGCCGTTATGC 3566
1213 ThrProGluProThrProThrProValProGluLysArgIleThrProSe 1229

3567 CAATAGCGGTTTGAGTGAATTTTCCGCCACGCTCAACAGCGTTTTC.... 3612
1229 rThrAlaAlaValLeuAsnMetAlaAlaThrLeuProLeuValPheAspA 1246

3613 ..........GCCGTACAGGACGAATTGGACCGCGTGTTTGCCGAAGAC 3651
1246 laGluLeuAsnSerIleArgGluArgLeuAsnIleMetLysAlaSerPro 1262

3652 CGCCGCAACGCCGTTTGGACAAGCGGCATCCGGGACACCAAACACTACCG 3701
1263 HisAsnAsnAsnValTrp...GlyAlaThrTyrAsnThrArgAsnAsnVa 1278

3702 TTCGCAAGATTTCCGCGCCTACCGCCAACAAACCGACCTGCGCCAAATCG 3751
1278 lThrThrAspAlaGlyAlaGlyPheGluGlnThr.....LeuThrG 1292

3752 GTATGCAGAAAAACCTC.....GGCAGCGGG 3777
1292 lyMetThrValGlyIleAspSerProAsnAspIleProGluGlyIleAla 1308

3778 CGCGTCGGCATCCTG......TTTTCGCACAACCGGACCGGAAACACCTT 3821
1309 ThrLeuGlyAlaPheMetGlyTyrSerHisSerHisIleGly......Ph 1323

3822 CGACGACGGCATCGGCAACTCGGCACGGCTTGCCCACGGTGCCGTTTTCG 3871
1323 eAspArgGly.....GlyHisGlySerVal...G 1332

3872 GGCAATACGGCATCGGC..... 3888
1332 lySerTyrSerLeuGlyGlyTyrAlaSerTrpGluHisGluSerGlyPhe 1348

3889 .....AGGTTCGACATCGGCATCAGC.. 3909
1349 TyrLeuAspGlyValValLysLeuAsnArgPheGluSerAsnValAlaGl 1365

3910 .......GCGGGCGCGGGTTTTAGTAGCGGCAGCCTT...TCAGACGGCA 3949
1365 yLysMetSerSerGlyGlyAlaAlaAsnGlySerTyrHisSerAsnGlyL 1382

3950 TCAGAGGCAAAATCCGCCGCCGCGTGCTGCATTACGGCATTCAGGCAAGA 3999
1382 euGlyGlyHis.....IleGluThrGly 1389

4000 TACCGCGCAGGTTTCGGCGGATTCGGCATCGAACCGCACATCGGCGCAAC 4049
1390 MetArgPheThrAspGlyAsnTrpAsnLeuThrProTyrAlaSerLeuTh 1406

4050 GCGCTATTTCGTCCAAAAAGCGGATTACCGA.....TACGAAA 4087
1406 rGlyPheThrAlaAspAsnProGluTyrHisLeuSerAsnGlyMetGluS 1423

4088 ACGTCAATATCGCCACCCCGGGCCTTGCATTCAACCGCTACCGC...GCG 4134
1423 erLysSerValAspThrArgSerIle.....TyrArgGluLeu 1435

4135 GGCATTAAGGCAGATTATTCATTCAAACCGGCGCAACACATTTCCATCAC 4184
1436 GlyAlaThrLeuSerTyrAsnMetArgLeuGlyAsnGlyMetGluIleGl 1452

4185 GCCTTATTTGAGCCTGTCCTATACCGATGCCGCTTCCGGCAAAGTCCGAA 4234
1452 uProTrpLeuLysAlaAlaValArgLysGluPheValAspAspAsnArgV 1469

4235 CGCGCGTCAATACC..... 4248
1469 alLysValAsnAsnAspGlyAsnPheValAsnAspLeuSerGlyArgArg 1485

4249 .....GCCGTATTGGCGCAGGATTTCGGCAAAACCCGCAGTGC 4286
1486 GlyIleTyrGlnAlaGlyIleLysAlaSerPheSerSerThrLeuSerGl 1502

4287 GGAATGGGGCGTAAACGCCGAAATCAAAGGTTTCACGCTGTCCCTCCACG 4336
1502 yHisLeuGlyVal.....GlyTyrSer.....HisG 1511

4337 CTGCCGCCGCCAAGGGGCCG 4356
1511 lyAlaGlyValGluSerPro 1517
```

seq_name: pir2:E95206

seq_documentation_block:
cell wall surface anchor family protein [imported] - Streptococcus pneumoniae (strain
C;Species: Streptococcus pneumoniae
C;Date: 03-Aug-2001 #sequence_revision 03-Aug-2001 #text_change 03-Aug-2001
C;Accession: E95206
R;Tettelin, H.; Nelson, K.E.; Paulsen, I.T.; Eisen, J.A.; Read, T.D.; Peterson, S.; H
on, J.D.; Umayam, L.A.; White, O.; Salzberg, S.L.; Lewis, M.R.; Radune, D.; Holtzappl
nson, T.; Hickey, E.K.; Holt, I.E.
Science 293, 498-506, 2001
A;Authors: Loftus, B.J.; Yang, F.; Smith, H.O.; Venter, J.C.; Dougherty, B.A.; Morris
A;Title: Complete Genome Sequence of a virulent isolate of Streptococcus pneumoniae.
A;Reference number: A95000; MUID:21357209; PMID:11463916
A;Accession: E95206
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-4776 <KUR>
A;Cross-references: GB:AE005672; PIDN:AAK75846.1; PID:g14973269; GSPDB:GN00164; TIGR:
A;Experimental source: strain TIGR4
C;Genetics:
A;Gene: SP1772

alignment_scores:
Quality: 329.50 Length: 994
Ratio: 0.586 Gaps: 27
Percent Similarity: 56.539 Percent Identity: 20.724

alignment_block:
US-09-303-518D-653 x E95206 ..

Align seg 1/1 to: E95206 from: 1 to: 4776

```
1501 TTTAGTGAAATCGGCTTGGTCAGCGGCAGGGGGACGGTGCAACTGAATGC 1550
 186 PheSerGlyAsnThrIleValAsnGlyAlaProAlaIleAsnAlaSerLe 202

1551 CGATAATCAGTTCAACCCCGACAAACTCTATTTCGGCTTTCGCGGCGGAC 1600
 202 uAsnIleAlaLysSerGluThrLysValTyr.....ThrGlyGluG 216

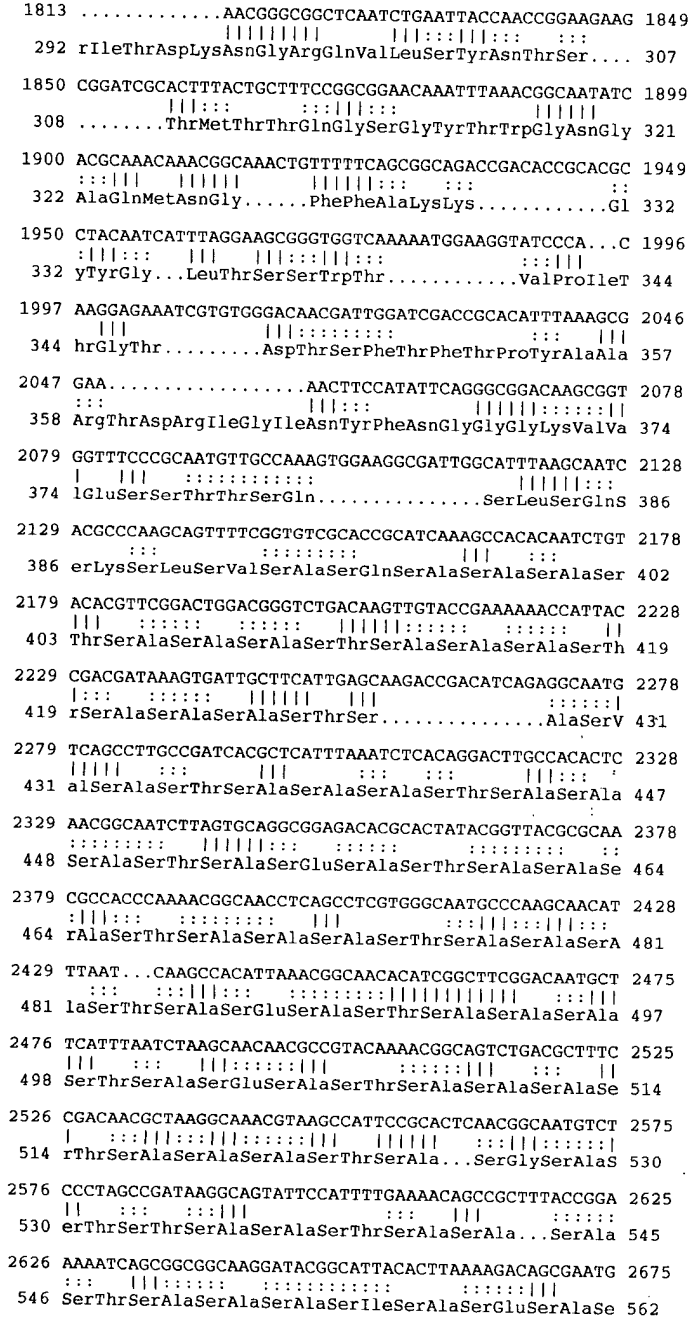
1601 GTTTGGATTTGAACGGGCATTCGCTTTCGTTCCACCGCATTCAAAATACC 1650
 216 lyValAspSerValTyrArgValProIleTyrTyrLysLeuLysValThr 232

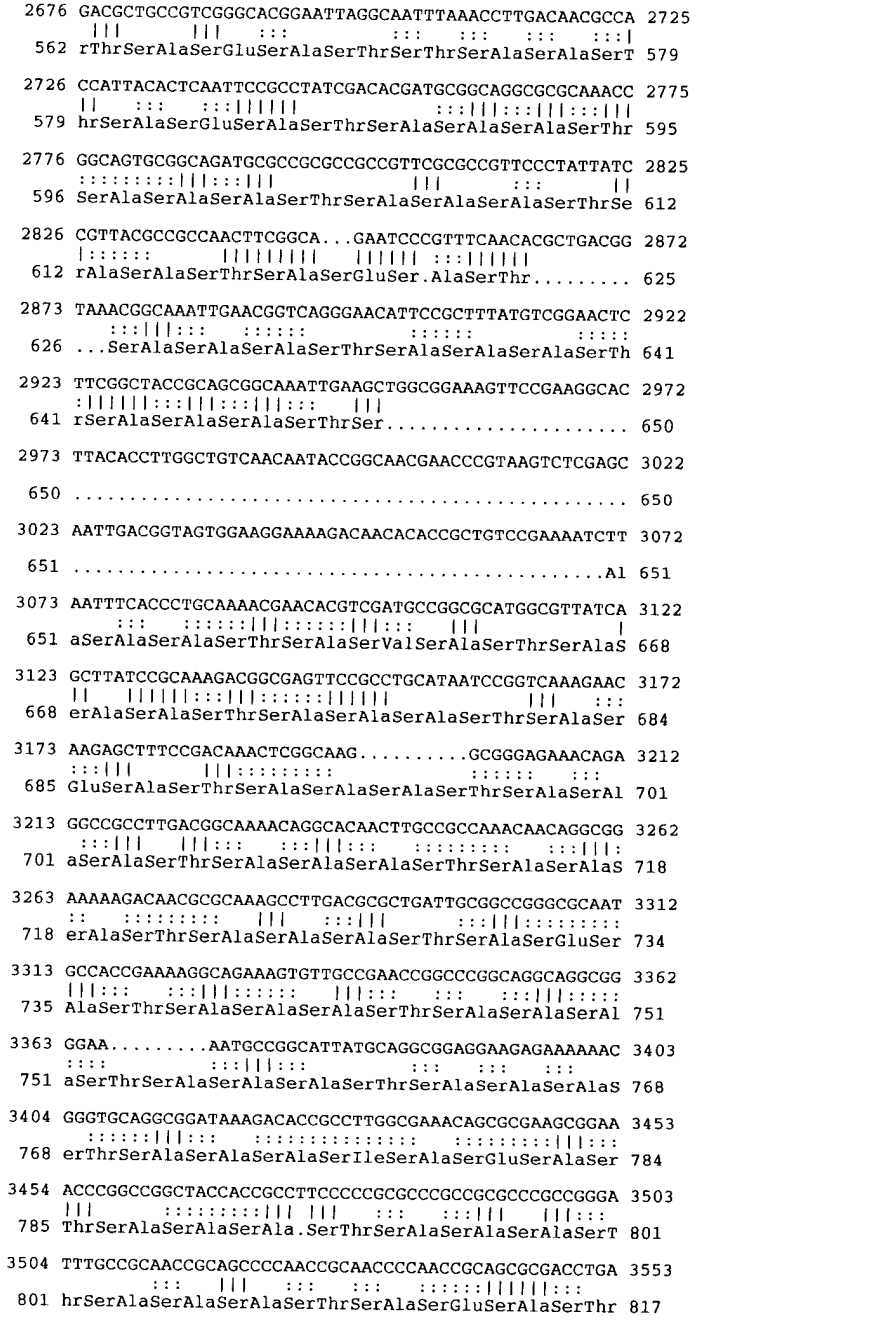
1651 GATGAAGGGGCGATGATTGTCAACCACAATCAAGACAAAGAATCCACCGT 1700
 233 AsnAspGlySerLysLeuThr.....PheThrTy 242

1701 TACCATTACA...GGCAATAAAGATATTACTACAACCGGCAAT..... 1740
 242 rThrValThrTyrValAsnProLysThrAsnAspLeuGlyAsnIleSerS 259

1741 .....AACAACAACTTGGATAGCAAAAAA 1764
 259 erMetArgProGlyTyrSerIleTyrAsnSerGlyThrSerThrGlnThr 275

1765 GAAATTGCCTACAACGGTTGGTTTGGCGAGAAAGATGCAACCAAAACG.. 1812
 276 MetLeuThrLeuGlySerAspLeuGlyLysProSerGlyValLysAsnTy 292
```


1813 .....AACGGGCGGCTCAATCTGAATTACCAACCGGAAGAAG 1849
292 rIleThrAspLysAsnGlyArgGlnValLeuSerTyrAsnThrSer.... 307
1850 CGGATCGCACTTTACTGCTTTCCGGCGGAACAAATTTAAACGGCAATATC 1899
308 .....ThrMetThrThrGlnGlySerGlyTyrThrTrpGlyAsnGly 321
1900 ACGCAAACAAACGGCAAACTGTTTTTCAGCGGCAGACCGACACCGCACGC 1949
322 AlaGlnMetAsnGly.....PhePheAlaLysLys.....Gl 332
1950 CTACAATCATTTAGGAAGCGGGTGGTCAAAAATGGAAGGTATCCCA...C 1996
332 yTyrGly...LeuThrSerSerTrpThr.....ValProIleT 344
1997 AAGGAGAAATCGTGTGGGACAACGATTGGATCGACCGCACATTTAAAGCG 2046
344 hrGlyThr.....AspThrSerPheThrPheThrProTyrAlaAla 357
2047 GAA.....AACTTCCATATTCAGGGCGGACAAGCGGT 2078
358 ArgThrAspArgIleGlyIleAsnTyrPheAsnGlyGlyGlyLysValVa 374
2079 GGTTTCCCGCAATGTTGCCAAAGTGGAAGGCGATTGGCATTTAAGCAATC 2128
374 lGluSerSerThrThrSerGln.....SerLeuSerGlnS 386
2129 ACGCCCAAGCAGTTTTCGGTGTCGCACCGCATCAAAGCCACACAATCTGT 2178
386 erLysSerLeuSerValSerAlaSerGlnSerAlaSerAlaSerAlaSer 402
2179 ACACGTTCGGACTGGACGGGTCTGACAAGTTGTACCGAAAAAACCATTAC 2228
403 ThrSerAlaSerAlaSerAlaSerThrSerAlaSerAlaSerAlaSerTh 419
2229 CGACGATAAAGTGATTGCTTCATTGAGCAAGACCGACATCAGAGGCAATG 2278
419 rSerAlaSerAlaSerAlaSerThrSer.....AlaSerV 431
2279 TCAGCCTTGCCGATCACGCTCATTTAAATCTCACAGGACTTGCCACACTC 2328
431 alSerAlaSerThrSerAlaSerAlaSerAlaSerThrSerAlaSerAla 447
2329 AACGGCAATCTTAGTGCAGGCGGAGACACGCACTATACGGTTACGCGCAA 2378
448 SerAlaSerThrSerAlaSerGluSerAlaSerThrSerAlaSerAlaSe 464
2379 CGCCACCCAAAACGGCAACCTCAGCCTCGTGGGCAATGCCCAAGCAACAT 2428
464 rAlaSerThrSerAlaSerAlaSerAlaSerThrSerAlaSerAlaSerA 481
2429 TTAAT...CAAGCCACATTAAACGGCAACACATCGGCTTCGGACAATGCT 2475
481 laSerThrSerAlaSerGluSerAlaSerThrSerAlaSerAlaSerAla 497
2476 TCATTTAATCTAAGCAACAACGCCGTACAAAACGGCAGTCTGACGCTTTC 2525
498 SerThrSerAlaSerGluSerAlaSerThrSerAlaSerAlaSerAlaSe 514
2526 CGACAACGCTAAGGCAAACGTAAGCCATTCCGCACTCAACGGCAATGTCT 2575
514 rThrSerAlaSerAlaSerAlaSerThrSerAla...SerGlySerAlaS 530
2576 CCCTAGCCGATAAGGCAGTATTCCATTTTGAAAACAGCCGCTTTACCGGA 2625
530 erThrSerThrSerAlaSerAlaSerThrSerAlaSerAla...SerAla 545
2626 AAAATCAGCGGCGGCAAGGATACGGCATTACACTTAAAAGACAGCGAATG 2675
546 SerThrSerAlaSerAlaSerAlaSerIleSerAlaSerGluSerAlaSe 562


2676 GACGCTGCCGTCGGGCACGGAATTAGGCAATTTAAACCTTGACAACGCCA 2725
562 rThrSerAlaSerGluSerAlaSerThrSerThrSerAlaSerAlaSerT 579
2726 CCATTACACTCAATTCCGCCTATCGACACGATGCGGCAGGCGCGCAAACC 2775
579 hrSerAlaSerGluSerAlaSerThrSerAlaSerAlaSerAlaSerThr 595
2776 GGCAGTGCGGCAGATGCGCCGCGCCGCCGTTCGCGCCGTTCCCTATTATC 2825
596 SerAlaSerAlaSerAlaSerThrSerAlaSerAlaSerAlaSerThrSe 612
2826 CGTTACGCCGCCAACTTCGGCA...GAATCCCGTTTCAACACGCTGACGG 2872
612 rAlaSerAlaSerThrSerAlaSerGluSer.AlaSerThr..... 625
2873 TAAACGGCAAATTGAACGGTCAGGGAACATTCCGCTTTATGTCGGAACTC 2922
626 ...SerAlaSerAlaSerAlaSerThrSerAlaSerAlaSerAlaSerTh 641
2923 TTCGGCTACCGCAGCGGCAAATTGAAGCTGGCGGAAAGTTCCGAAGGCAC 2972
641 rSerAlaSerAlaSerAlaSerThrSer..... 650
2973 TTACACCTTGGCTGTCAACAATACCGGCAACGAACCCGTAAGTCTCGAGC 3022
650 650
3023 AATTGACGGTAGTGGAAGGAAAAGACAACACACCGCTGTCCGAAAATCTT 3072
651 .....Al 651
3073 AATTTCACCCTGCAAAACGAACACGTCGATGCCGGCGCATGGCGTTATCA 3122
651 aSerAlaSerAlaSerThrSerAlaSerValSerAlaSerThrSerAlaS 668
3123 GCTTATCCGCAAAGACGGCGAGTTCCGCCTGCATAATCCGGTCAAAGAAC 3172
668 erAlaSerAlaSerThrSerAlaSerAlaSerAlaSerThrSerAlaSer 684
3173 AAGAGCTTTCCGACAAACTCGGCAAG.....GCGGGAGAAACAGA 3212
685 GluSerAlaSerThrSerAlaSerAlaSerAlaSerThrSerAlaSerAl 701
3213 GGCCGCCTTGACGGCAAAACAGGCACAACTTGCCGCCAAACAACAGGCGG 3262
701 aSerAlaSerThrSerAlaSerAlaSerAlaSerThrSerAlaSerAlaS 718
3263 AAAAAGACAACGCGCAAAGCCTTGACGCGCTGATTGCGGCCGGGCGCAAT 3312
718 erAlaSerThrSerAlaSerAlaSerAlaSerThrSerAlaSerGluSer 734
3313 GCCACCGAAAAGGCAGAAAGTGTTGCCGAACCGGCCCGGCAGGCAGGCGG 3362
735 AlaSerThrSerAlaSerAlaSerAlaSerThrSerAlaSerAlaSerAl 751
3363 GGAA.....AATGCCGGCATTATGCAGGCGGAGGAAGAGAAAAAAC 3403
751 aSerThrSerAlaSerAlaSerAlaSerThrSerAlaSerAlaSerAlaS 768
3404 GGGTGCAGGCGGATAAAGACACCGCCTTGGCGAAACAGCGCGAAGCGGAA 3453
768 erThrSerAlaSerAlaSerAlaSerIleSerAlaSerGluSerAlaSer 784
3454 ACCCGGCCGGCTACCACCGCCTTCCCCCGCGCCCGCCGCGCCCCGCCGGGA 3503
785 ThrSerAlaSerAlaSerAla.SerThrSerAlaSerAlaSerAlaSerT 801
3504 TTTGCCGCAACCGCAGCCCCAACCGCAACCCCAACCGCAGCGCGACCTGA 3553
801 hrSerAlaSerAlaSerAlaSerThrSerAlaSerGluSerAlaSerThr 817
3554 TCAGCCGTTATGCCAATAGCGGTTTGAGTGAATTTTCCGCCACGCTCAAC 3603

```
      |||||||             ::::::   ::::::             :::::
  818 SerAlaSerAlaSerAlaSerThrSerAlaSerAlaSerAlaSerThrSe 834

 3604 AGCGTTTTCGCCGTACAGGACGAATTGGACCGCGTGTTTGCCGAAGACCG 3653
      :|||  :::         |||:::  :::|||          :::|
  834 rAlaSerAlaSerAlaSerThrSerAlaSerAlaSerAlaSerThrSerA 851

 3654 CCGCAACGCCGTTTGGACAAGCGGCATCCGGGACACCAAACACTACCGTT 3703
      ||::::::     :::   :::|||||||:::::   :::|||:::::::
  851 laSerAlaSerAlaSerThrSerAlaSerAlaSerAlaSerThrSerAla 867

 3704 CGCAAGATTTCCGCGCCTACCGCCAACAAACCGACCTGCGCCAAATCGGT 3753
         :::   ::::::   :::|||:::      :::   :::::::::
  868 SerGluSerAlaSerThrSerAlaSerAlaSerAlaSerThrSerAlaSe 884

 3754 ATGCAGAAAAACCTCGGCAGCGGGCGCGTCGGCATCCTGTTTTCGCACAA 3803
                :::::::::|||:::||||||||||||          |||:
  884 rGluSerAlaSerThrSerAlaSerAlaSerAlaSer.........ThrS 898

 3804 CCGGACCGGAAACACCTTCGACGACGGCATCGGCAACTCGGCACGGCTTG 3853
      :::::    :::   ||||||:::||||||||:::
  898 erAlaSerAlaSerAlaSerThrSerAlaSerAlaSerAlaSerThrSer 914

 3854 CCCACGGTGCCGTTTTCGGGCAATACGGCATCGGCAGGTTCGACATCGGC 3903
         ::::::      |||   ::::::::::::::::::||||||||||
  915 AlaSerAlaSerAlaSerThrSerAlaSerAlaSerAlaSerThrSerAl 931

 3904 ATCAGCGCGGGCGCGGGTTTTAGTAGCGGCAGCCTTTCAGACGGCATCAG 3953
      |||||||   |||  :::  :::::::|||:::  ::::::::::::
  931 aSerAlaSerAlaSerThrSerAlaSerAlaSerThrSerAlaSerGluS 948

 3954 AGGCAAAATCCGCCGCCGCGTGCTGCATTACGGCATTCAGGCAAGATACC 4003
      ::|||::::::::::|||::::::      ::::::      :::   :::
  948 erAlaSerThrSerAlaSerAlaSerAlaSerThrSerAlaSerAlaSer 964

 4004 GCGCAGGTTTCGGCGGATTCGGCATCGAACCGCACATCGGCGCAACGCGC 4053
      |||::::::||||||:::::::::::::::::  ::::::::::::  ||
  965 AlaSerThrSerAlaSerAlaSerAlaSerThrSerAlaSerGluSerAl 981

 4054 TATTTCGTCCAAAAAGCGGATTACCGATACGAAAACGTCAATATCGCCAC 4103
      |  :::|||  :::      :::::::|||:::::|||  |||
  981 aSerThrSerAlaSerAlaSerAlaSerThrSerAlaSerAlaSerAlaS 998

 4104 CCCGGGCCTTGCATTCAACCGCTACCGCGCGGGCATTAAGGCAGATTATT 4153
         :::      ::::::|||:::::::   |||  :::|||:::   |
  998 erThrSerAlaSerAlaSerAlaSerThrSerAla..SerAlaSerAlaS 1014

 4154 CATTCAAACCGGCGCAACACATTTCCATCACGCCTTATTTGAGCCTGTCC 4203
      ||  :::  :::       |||||||::       |||  |||
 1014 erThrSerAlaSerAlaSerAlaSerIleSerAlaSerGluSerAlaSer 1030

 4204 TATACCGATGCCGCTTCCGGCAAAGTCCGAACGCGCGTCAATACCGCCGT 4253
         :::  :::|||||:::::::::  :::  |||:::::::::::
 1031 ThrSerAlaSerAlaSerAlaSerThrSerAlaSerValSerAlaSerTh 1047

 4254 ATTGGCGCAGGATTTCGGCAAAACCCGCAGTGCGGAATGGGGCGTAAACG 4303
      :  |||:::     ::::::|||  ::::::|||  :::  ::::
 1047 rSerAlaSerAlaSerAlaSerThrSerAlaSerGluSerAlaSerThrS 1064

 4304 CCGAAATCAAAGGTTTCACGCTGTCCCTCCACGCTGCCGCCGCCAAGGGG 4353
      ::       :::   :::   |||        :::|||:::   :::
 1064 erAlaSerAlaSerAlaSerThrSerAlaSerGluSerAlaSerThrSer 1080

 4354 CCGCAATTGGAAGCGCAGCACAGCGCG 4380
         :::   :::|||:::   ||||||
 1081 AlaSerAlaSerAlaSerThrSerAla 1089

seq_name: pir2:T08658

seq_documentation_block:
hypothetical protein f949 - Escherichia coli
C;Species: Escherichia coli
C;Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 11-Jun-1999
C;Accession: T08658
R;Plunkett, G.
submitted to the EMBL Data Library, September 1995
A;Reference number: Z16465
A;Accession: T08658
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-949 <PLU>
A;Cross-references: EMBL:U36840; NID:g1033110; PID:g1033141
A;Experimental source: strain k12, substrain MG1655
C;Genetics:
A;Map position: 57 min

alignment_scores:
            Quality:  295.50              Length:   1157
              Ratio:  0.554                 Gaps:     57
 Percent Similarity:  46.067   Percent Identity: 20.138

alignment_block:
 US-09-303-518D-653 x T08658   ..

 Align seg 1/1  to: T08658  from: 1  to: 949

 1420 GAAAACCAAGGCTCGGTCAGCGTGGGCGACGGTAAAGTCATCTTAGATCA 1469
      ||||||   ||||||::::::|||      |||:::      :::|||
    3 GluAsnAlaGlySerLeuThrValValThrGlySerArgAlaValAspTh 19

 1470 GCAGGCGGACGATCAAGGCAAAAAACAAGCCTTT...AGTGAAATCGGCT 1516
             :::   :::||||||   :::::::::   ::::::::::|||
   19 rIleIleAsnAlaAsnGlyLysMetAspValTyrGlyLysAspValGlyT 36

 1517 TGGTC.........AGCGGCAGGGGGACGGTGCAACTGAATGCCGATAAT 1557
        |||         :::|||      |||:::      :::|||   :::
   36 hrValLeuAsnSerAlaGlyThrGlnThrIleTyrAlaSerAlaThrSer 52

 1558 CAGTTCAACCCCGACAAACTCTATTTCGGCTTTCGCGGCGGACGTTTGGA 1607
                 ||||||         :::::::::|||||||:::
   53 ............AspLysAla......AsnIleLysGlyGlyLysGlnTh 63

 1608 TTTGAACGGGCATTCGCTTTCGTTCCACCGCATTCAAAATACCGATGAAG 1657
       :::   |||   :::  :::                  |||   :::::|
   63 rValTyrGlyLeuAlaThrGlu............AlaAsnIleGluSerG 76

 1658 GGGCGATGATTGTCAACCACAATCAAGACAAAGAATCCACCGTTACCATT 1707
      ||   :::|||||||:::   ::::::   :::::::::   ::::::
   76 lyGluGlnIleValAspGlyGlySerThrGluLysThrHisIleAsn... 91

 1708 ACAGGCAATAAAGATATTACTACAACCGGCAATAACAACAACTTGGAT.. 1755
         |||::::::   :::   :::   |||:::      |||   |||
   92 GlyGlyThrGlnThrValGlnAsnTyrGlyLysAlaIleAsnThrAspIl 108

 1756 .......AGCAAAAAAGAAATTGCCTACAACGGTTGGTTTGGCGAGAAAG 1798
             :::   ::::::|||      |||         |||      :
  108 eValSerGlyLeuGlnGlnIleMetAlaAsn.........GlyThrAlaG 122

 1799 ATGCAACCAAAACGAACGGGCGGCTCAATCTGAATTACCAACCGGAAGAA 1848
      ::::::::      ||||||      ::::::      :::
  122 luGlySerIleIleAsnGlyGlySerGlnValValAsnGluGlyGlyLeu 138

 1849 GCGGATCGCACTTTACTGCTTTCCGGCGGAACA..................... 1881
      |||:::::::::::::|||   :::|||||||||
  139 AlaGluAsnSerValLeuAsnAspGlyGlyThrLeuAspValArgGluLy 155

 1882 ....AATTTAAACGGCAATATCACGCAAACAAACGGCAAACTGTTTTTCA 1927
          :::   :::|||         :::::::::|||   |||      :
  155 sGlySerAlaThrGlyIleGlnGlnSerSerGlnGlyAlaLeuValAlaT 172

 1928 GCGGCAGACCGACACCGCACGCCTACAATCATTTAGGAAGCGGGTGGTCA 1977
      ::   |||   |||       :::                       :::
  172 hrThrArgAlaThrArgValThr.....................GlyThr 181

 1978 AAAATGGAAGGTATCCCACAAGGAGAAATCGTGTGGGACAACGATTGGAT 2027
```


182 ArgAlaAspGlyValAla 187
2028 CGACCGCACATTTAAAGCGGAAAACTTCCATATTCAGGGCGGACAAGCG. 2076
188 PheSerIleGluGlnGlyAlaAlaA 196
2077 GTGGTTTCCCGCAATGTTGCCAAAGTGGAAGGCGATTGG 2115
196 snAsnIleLeuLeuAlaAsnGlyGlyValLeuThrValGluSerAspThr 212
2116 CATTTAAGCAATCACGCCCAAGCAGTTTTCGGTGTCGCACCGCATCAAAG 2165
213 Ser...SerAspLysThrGlnValAsnMetGly GlyAr 224
2166 CCACACAATCTGTACACGTTCGGACTGGACGGGTCTGACAAGTTGTACCG 2215
224 gGluIleValLysThrLysAlaThrAlaThrGlyThrThrLeuThrGlyG 241
2216 AAAAAACCATTACCGACGATAAAGTGATTGCTTCATTGAGCAAGACCGAC 2265
241 lyGluGlnIleValGlu 246
2266 ATCAGAGGCAATGTCAGCCTTGCCGATCACGCTCATTTAAATCTCACAGG 2315
247 GlyValAlaAsnGluThrThrIleAsnAspGlyGl 258
2316 ACTTGCCACACTCAACGGCAAT CTTAGTG 2344
258 yIleGlnThrValSerAlaAsnGlyGluAlaIleLysThrLysIleAsnG 275
2345 CAGGCGGAGACACGCACTATACGGTTACGCGCAACGCCACCCAAAACGGC 2394
275 luGlyGly ThrLeuThrValAsn AspAsnGly 285
2395 AACCTCAGC...CTCGTGGGCAATGCCCAAGCAACATTTAATCAAGCCAC 2441
286 LysAlaThrAspIleValGlnAsnSerGlyAlaAlaLeuGlnThrSerTh 302
2442 ATTAAACGGC AACACATCGGCTTCGGACAATGCTTCATTTAATC 2485
302 rAlaAsnGlyIleGluIleSerGlyThrHisGlnTyrGlyThrPheSerI 319
2486 TAAGCAACAACGCCGTACAAAAC GGCAGTCTG 2517
319 leSerGlyAsnLeuAlaThrAsnMetLeuLeuGluAsnGlyGlyAsnLeu 335
2518 ACGCTTTCCGACAACGCTAAGGCAAACGTAAGCCATTCCGCACTCAACGG 2567
336 LeuValLeuAlaGlyThrGluAlaArgAspSerThrValGlyLysGlyGl 352
2568 CAATGTCTCCCTAGCCGATAAGGCAGTATTCCATTTTGAAAACAGCCGCT 2617
352 y AlaMetGlnAsnLeuGlyGlnAspSerA 362
2618 TTACCGGAAAAATCAGCGGCGGCAAGGATACGGCATTACACTTAAAAGAC 2667
362 laThrLysValAsnSerGlyGlyGlnTyrThrLeuGlyArgSerLysAsp 378
2668 AGCGAATGGACGCTGCCGTCGGGCACGGAATTA 2700
379 GluPheGlnAlaLeuAlaArgAlaGluAspLeuGlnValAlaGlyGlyTh 395
2700 2700
395 rAlaIleValTyrAlaGlyThrLeuAlaAspAlaSerValSerGlyAlaT 412
2701 ..GGCAATTTAAACCTT 2715
412 hrGlySerLeuSerLeuMetThrProArgAspAsnValThrProValLys 428
2716 GACAACGCCACCATTACACTCAATTC 2741
429 LeuGluGlyAlaValArgIleThrAspSerAlaThrLeuThrLeuGlyAs 445
2742 CGCCTAT CGACACGATGCGGCAGGCGCGCAAACCGGCAGT. 2781
445 nGlyValAspThrThrLeuAlaAspLeuThrAlaAlaSerArgGlySerV 462
2782 GCGGCAGATGCGCCGCGCCGCCGT 2805
462 alTrpLeuAsnSerAsnAsnSerCysAlaGlyThrSerAsnCysGluTyr 478
2806 TCGCGCCGTTCCCTATTA TC 2825
479 ArgValAsnSerLeuLeuLeuAsnAspGlyAspValTyrLeuSerAlaGl 495
2826 CGTTACGCCGCCAACTTCGGCAGAATCCCGTTTCAACACGCTGACGGTAA 2875
495 nThrAlaAlaProAlaThrThrAsnGlyIleTyrAsnThrLeuThrThrA 512
2876 ACGGCAAATTGAACGGTCAGGGAACATTCCGCTTTATGTCGGAACTCTTC 2925
512 sn...GluLeuSerGlySerGlyAsnPheTyrLeuHisThrAsnValAla 527
2926 GGCTACCGCAGCGGCAAATTGAAGCTGGCGGAAAGTTCCGAAGGCACTTA 2975
528 GlySerArgGlyAspGlnLeuValValAsnAsnAsnAlaThrGlyAsnPh 544
2976 CACCTTGGCTGTCAACAATACCGGCAACGAACCCGTAAGTCTCGAGCAAT 3025
544 eLysIlePheValGlnAspThrGlyValSerProGlnSerAspAspAlaM 561
3026 TGACGGTAGTGGAA GGAAAAGACAACACACCGCTGTCCGAAAAT 3069
561 etThrLeuValLysThrGlyGlyGlyAsp 570
3070 CTTAATTTCACCCTGCAAAACGAA CACGTCGATGCCGGCGCATG 3113
571 AlaSerPheThrLeuGlyAsnThrGlyGlyPheValAspLeuGlyThrTy 587
3114 GCGTTATCAGCTTATCCGCAAAGACGGCGAGTTCCGCCTGCATAATCCGG 3163
587 rGluTyr...ValLeuLysSerAspGly.AsnSer 597
3164 TCAAAGAACAAGAGCTTTCCGACAAACTCGGCAAGGCGGGAGAAACAGAG 3213
597 597
3214 GCCGCCTTGACGGCAAAACAGGCACAACTTGCCGCCAAACAACAGGCGGA 3263
597 597
3264 AAAAGACAACGCGCAAAGCCTTGACGCGCTGATTGCGGCCGGGCGCAATG 3313
598 AsnTrpAsnLeuThrAsnAspValLysProAsnProAsp 610
3314 CCACCGAAAAGGCAGAAAGTGTTGCCGAACCGGCCCGGCAGGCAGGCGGG 3363
611 Pro IleProAsn 614
3364 GAAAATGCCGGCATTATGCAGGCGGAGGAAGAGAAAAAACGGGTGCAGGC 3413
614 614
3414 GGATAAAGACACCGCCTTGGCGAAACAGCGCGAAGCGGAAACCCGGCCGG 3463
615 ProLysProAspPro 619
3464 CTACCACCGCCTTCCCCCGCGCCCGCCGCGCCCGCCGGGATTTGCCGCAA 3513
620 LysProAspProLysProAspPro..AsnProLysProAsp...ProThr 634
3514 CCGCAGCCCCAACCGCAACCCCAACCGCAGCGCGACCTGATCAGCCGTTA 3563
635 ProAspProThrProThrProValProGluLysArgIleThrPro 649

```
3564 TGCCAATAGCGGTTTGAGTGAATTTTCCGCCACGCTCAACAGCGTTTTC. 3612
     :::::::::::::::   :::::::::|||||||||      ||||||
 650 .SerThrAlaAlaValLeuAsnMetAlaAlaThrLeuProLeuValPheA 666

3613 .............GCCGTACAGGACGAATTGGACCGCGTGTTTGCCGAA 3648
                 ::::::::::::::|||:::   :::   |||:::
 666 spAlaGluLeuAsnSerIleArgGluArgLeuAsnIleMetLysAlaSer 682

3649 GACCGCCGCAACGCCGTTTGGACAAGCGGCATCCGGGACACCAAACACTA 3698
       ::::::|||   ||||||   ::::::      :::|||:::::::
 683 ProHisAsnAsnAsnValTrp...GlyAlaThrTyrAsnThrArgAsnAs 698

3699 CCGTTCGCAAGATTTCCGCGCCTACCGCCAACAAACCGACCTGCGCCAAA 3748
       :::   |||      |||
 698 nValThrThrAspAlaGlyAla........................... 705

3749 TCGGTATGCAGAAAAACCTCGGCAGCGGGCGCGTCGGCATC....... 3789
     |||:::::::::::::|||   :::      |||||||||
 706 ..GlyPheGluGlnThrLeuThrGlyMetThrValGlyIleAsp***Arg 721

3790 .....................................CTGTTTTC 3797
                                          :::||
 722 AsnAspIleProGluGlyIleThrThrLeuGlyAlaPheMetGlyTyrSe 738

3798 GCACAACCGGACCGGAAACACCTTCGACGACGGCATCGGCAACTCGGCAC 3847
     ||||::::::   |||      ||||||   |||
 738 rHisSerHisIleGly......PheAspArgGly.............. 747

3848 GGCTTGCCCACGGTGCCGTTTTCGGGCAATACGGCATCGGC......... 3888
       :::||||||:::|||   |||:::|||:::::::|||
 748 .....GlyHisGlySerVal...GlySerTyrSerLeuGlyGlyTyrAla 761

3888 ................................................. 3888

 762 SerTrpGluHisGluSerGlyPheTyrLeuAspGlyValValLysLeuAs 778

3889 .AGGTTCGACATCGGCATCAGC.........GCGGGCGCGGGTTTTAGTA 3928
      ||||||      :::::::::         ::::::::|||   ::::
 778 nArgPheLysSerAsnValAlaGlyLysMetSerSerGlyGlyAlaAlaA 795

3929 GCGGCAGCCTT...TCAGACGGCATCAGAGGCAAAATCCGCCGCCGCGTG 3975
     ::|||||||      |||:::|||:::   |||
 795 snGlySerTyrHisSerAsnGlyLeuGlyGlyHis.............. 806

3976 CTGCATTACGGCATTCAGGCAAGATACCGCGCAGGTTTCGGCGGATTCGG 4025
                |||::::::      |||         |||:::::::::
 807 ............IleGluThrGlyMetArgPheThrAspGlyAsnTrpAs 819

4026 CATCGAACCGCACATCGGCGCAACGCGCTATTTCGTCCAAAAAGCGGATT 4075
     ::::   |||:::   :::   |||   :::   :::::::::   :::|
 819 nLeuThrProTyrAlaSerLeuThrGlyPheThrAlaAspAsnProGluT 836

4076 ACCGATACGAAAAC............GTCAATATCGCCACCCCGGGCCTT 4113
     ||:::   :::|||                ::::::   |||   ::::::
 836 yrHisLeuSerAsnGlyMetLysSerLysSerValAspThrArgSerIle 852

4114 GCATTCAACCGCTACCGC...GCGGGCATTAAGGCAGATTATTCATTCAA 4160
                |||||      |||         :::|||:::::::::
 853 ............TyrArgGluLeuGlyAlaThrLeuSerTyrAsnMetAr 865

4161 ACCGGCGCAACACATTTCCATCACGCCTTATTTGAGCCTGTCCTATACCG 4210
     :   ::::::   :::::::::   |||:::|||:::   :::
 865 gLeuGlyAsnGlyMetGluValGluProTrpLeuLysAlaAlaValArgL 882

4211 ATGCCGCTTCCGGCAAAGTCCGAACGCGCGTCAATACCGCCGTATTGGCG 4260
                          |||::::::|||||||:::
 882 ysGluPheValAspAspAsnArgValLysValAsnSerAspGlyAsnPhe 898

4261 CAGGATTTC......GGCAAAACCCGCAGTGCGGAATGGGGCGTAAACGC 4304
       ::::::      |||:::           :::   |||:::::::||
 899 ValAsnTyrLeuSerGlyArgArgGlyIleTyrGlnAlaGlyIleLysAl 915

4305 CGAAATCAAAGGTTTCACGCTGTCCCTCCACGCTGCCGCCGCCAAGGGGC 4354
     |:::::::::::   |||||||||   |||   :::::::::
  915 aSerPheSerSer...ThrLeuSerGlyHisLeuGlyValGly....... 928

4355 CGCAATTGGAAGCGCAGCACAGCGCGGGCATCAAA.............. 4389
                    :::||||||||||||:::::::
  929 ..........TyrSerHisSerAlaGlyValGluSerProTrpAsnAla 941

4390 ...TTAGGCTACCGCTGG 4404
           |||   :::|||
  942 ValAlaGlyValAsnTrp 947
```

seq_name: pir2:G64993

seq_documentation_block:
yfaL protein - Escherichia coli (strain K-12)
C;Species: Escherichia coli
C;Date: 12-Sep-1997 #sequence_revision 17-Sep-1997 #text_change 08-Oct-1999
C;Accession: G64993
R;Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; .A.; Rose, D.J.; Mau, B.; Shao, Y.
Science 277, 1453-1462, 1997
A;Title: The complete genome sequence of Escherichia coli K-12.
A;Reference number: A64720; MUID:97426617
A;Accession: G64993
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-1250 <BLAT>
A;Cross-references: GB:AE000313; GB:U00096; NID:g2367132; PIDN:AAC75293.1; PID:g1788
A;Experimental source: strain K-12, substrain MG1655
C;Genetics:
A;Gene: yfaL

alignment_scores:
Quality: 292.50 Length: 1466
Ratio: 0.447 Gaps: 72
Percent Similarity: 44.679 Percent Identity: 19.986

alignment_block:
US-09-303-518D-653 x G64993 ..

Align seg 1/1 to: G64993 from: 1 to: 1250

```
 637 AACCGCGAAAGTTCATATCATATTGCAAGCGCATATTCT.......... 675
     :::|||:::|||         |||::::::   :::|||
  46 SerArgGlnSerLeuSerGlyIleThrGlnAspTrpSerIleAlaAspGl 62

 676 ....TGGCTCGTCGGTGGCAATACCTTTGCACAAAATGGATCAGGTGGTG 721
         |||||||||   ::::::   :::::::::|||:::||||||||:
  62 yGlnTrpLeuValPheSerAsp...MetThrAsnAsnAlaSerGlyGlyA 78

 722 GCACAGTCAACTTAGGTAGCGAAAAAATTAAACATAGCCCATATGGTTTT 771
     :::::                                        |||
  78 laVal.....................................Phe 80

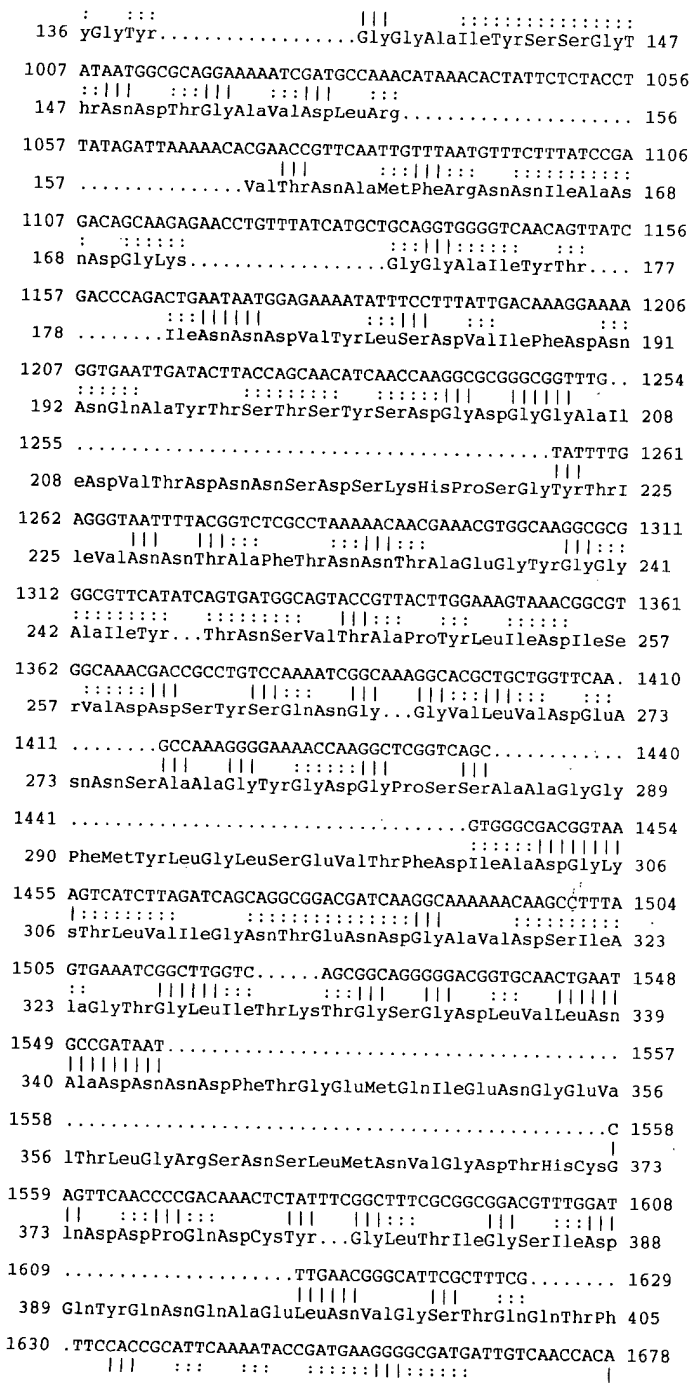
 772 TTACCAACAGGAGGCTCATTT..............GGCGACAGTGGCTC 806
     |||      |||::::::|||               ::::::::|||
  81 LeuGlnGlnGlyAlaGluPheSerLeuLeuProGluAsnGluThrGlyMe 97

 807 ACCAATGTTTATCTATGATGCCCAAAAGCAAAAGTGGTTAATTAATGGGG 856
         :::|||      ::::::                 ::::::|||
  97 tThrLeuPheAlaAsnAsnThr.................ValThrGlyG 108

 857 TATTGCAAACAGGCAACCCCTATATAGGAAAAAGCAATGGCTTCCAGCTA 906
           ::::::|||:::      ::::::|||:::|||:::
 108 luTyrAsnAsnGlyGlyAlaIlePheAlaLysGluAsnSerThrLeuAsn 124

 907 GTTCGTAAAGATTGGTTCTATGATGAAATCTTTGCTGGAGATACCCATTC 956
     :::                  |||   ||||||:::|||::::::   ::
 125 LeuThr..............AspValIlePheSerGlyAsnValAlaGl 136

 957 AGTATTCTACGAACCACATCAAAATGGGAAATACTTTTTTAACGACAATA 1006
```


136 yGlyTyr.....GlyGlyAlaIleTyrSerSerGlyT 147
1007 ATAATGGCGCAGGAAAAATCGATGCCAAACATAAACACTATTCTCTACCT 1056
147 hrAsnAspThrGlyAlaValAspLeuArg..... 156
1057 TATAGATTAAAAACACGAACCGTTCAATTGTTTAATGTTTCTTTATCCGA 1106
157 .....ValThrAsnAlaMetPheArgAsnAsnIleAlaAs 168
1107 GACAGCAAGAGAACCTGTTTATCATGCTGCAGGTGGGGTCAACAGTTATC 1156
168 nAspGlyLys.....GlyGlyAlaIleTyrThr.... 177
1157 GACCCAGACTGAATAATGGAGAAAATATTTCCTTTATTGACAAAGGAAAA 1206
178 .....IleAsnAsnAspValTyrLeuSerAspValIlePheAspAsn 191
1207 GGTGAATTGATACTTACCAGCAACATCAACCAAGGCGCGGGCGGTTTG.. 1254
192 AsnGlnAlaTyrThrSerThrSerTyrSerAspGlyAspGlyGlyAlaIl 208
1255 .....TATTTTG 1261
208 eAspValThrAspAsnAsnSerAspSerLysHisProSerGlyTyrThrI 225
1262 AGGGTAATTTTACGGTCTCGCCTAAAAACAACGAAACGTGGCAAGGCGCG 1311
225 leValAsnAsnThrAlaPheThrAsnAsnThrAlaGluGlyTyrGlyGly 241
1312 GGCGTTCATATCAGTGATGGCAGTACCGTTACTTGGAAAGTAAACGGCGT 1361
242 AlaIleTyr...ThrAsnSerValThrAlaProTyrLeuIleAspIleSe 257
1362 GGCAAACGACCGCCTGTCCAAAATCGGCAAAGGCACGCTGCTGGTTCAA. 1410
257 rValAspAspSerTyrSerGlnAsnGly...GlyValLeuValAspGluA 273
1411 .....GCCAAAGGGGAAAACCAAGGCTCGGTCAGC..... 1440
273 snAsnSerAlaAlaGlyTyrGlyAspGlyProSerSerAlaAlaGlyGly 289
1441 .....GTGGGCGACGGTAA 1454
290 PheMetTyrLeuGlyLeuSerGluValThrPheAspIleAlaAspGlyLy 306
1455 AGTCATCTTAGATCAGCAGGCGGACGATCAAGGCAAAAAACAAGCCTTTA 1504
306 sThrLeuValIleGlyAsnThrGluAsnAspGlyAlaValAspSerIleA 323
1505 GTGAAATCGGCTTGGTC......AGCGGCAGGGGGACGGTGCAACTGAAT 1548
323 laGlyThrGlyLeuIleThrLysThrGlySerGlyAspLeuValLeuAsn 339
1549 GCCGATAAT..... 1557
340 AlaAspAsnAsnAspPheThrGlyGluMetGlnIleGluAsnGlyGluVa 356
1558 .....C 1558
356 lThrLeuGlyArgSerAsnSerLeuMetAsnValGlyAspThrHisCysG 373
1559 AGTTCAACCCCGACAAACTCTATTTCGGCTTTCGCGGCGGACGTTTGGAT 1608
373 lnAspAspProGlnAspCysTyr...GlyLeuThrIleGlySerIleAsp 388
1609 .....TTGAACGGGCATTCGCTTTCG..... 1629
389 GlnTyrGlnAsnGlnAlaGluLeuAsnValGlySerThrGlnGlnThrPh 405
1630 .TTCCACCGCATTCAAAATACCGATGAAGGGGCGATGATTGTCAACCACA 1678
405 eValHisAlaLeuThrGlyPheGlnAsnGlyThrLeu.....A 418
1679 ATCAAGACAAAGAATCCACCGTTACC..... 1704
418 snIleAspAlaGlyGlyAsnValThrValAsnGlnGlySerPheAlaGly 434
1705 ...ATTACAGGCAATAAAGATATTACTACAACCGGCAATAACAACAAC.. 1749
435 IleIleGluGlyAlaGlyGlnLeuThrIleAlaGlnAsnGlySerTyrVa 451
1750 .TTGGATAGCAAAAAAGAAATTGCCTACAACGGTTGGTTTGGCGAGAAAG 1798
451 lLeuAlaGlyAlaGlnSerMetAlaLeuThrGly.....A 463
1799 ATGCAACCAAAACGAACGGGCGGCTCAATCTGAATTACCAACCGGAAGAA 1848
463 spIleValValAspAspGlyAlaVal...LeuSerLeuGluGlyAspAla 478
1849 GCGGATCGCACTTTACTGCTTTCCGGCGGAACAAAT.....TTAAACGG 1892
479 AlaAspLeuThrAlaLeuGlnAspAspProGlnSerIleValLeuAsnGl 495
1893 CAATATCACGCAAACAAACGGCAAACTGTTTTTCAGCGGCAGACCGACAC 1942
495 yGlyValLeuAspLeuSerAsp..... 502
1943 CGCACGCCTACAATCATTTAGGAAGCGGGTGGTCAAAAATGGAAGGTATC 1992
503 .....PheSerThrTrpGlnSerGlyThrSerTyrAsnAspGlyLeu 516
1993 .....CCACAAGGAGAAATCGTGTGGGACAACGATTGGATCGA 2030
517 GluValSerGlySerSerGlyThrValIleGlySerGlnAspValValAs 533
2031 CCGCACATTTAAAGCGGAAAACTTCCATATTCAGGGC...GGACAAGCGG 2077
533 pLeuAla...GlyGlyAspAsnLeuHisIleGlyGlyAspGlyLysAspG 549
2078 TGGTTTCCCGCAATGTTGCCAAAGTGGAAGGCGATTGGCATTTAAGCAAT 2127
549 lyValTyrValValValAspAlaSerAspGlyGlnValSerLeuAlaAsn 565
2128 CACGCCCAAGCAGTTTTCGGTGTCGCACCGCATCAAAGCCACACAATC.. 2175
566 Asn...AsnSerTyrLeuGlyThrThrGlnIleAlaSerGlyThrLeuMe 581
2176 .TGTACACGTTCGGACTGGACGGGTCTGACAAGTTGTACCGAAAAAACCA 2224
581 tValSerAspAsnSerGlnLeuGlyAspThrHisTyrAsnArgGlnValI 598
2225 TTACCGACGATAAA.....GTGATTGCTTCATTGAGCAAGACC 2262
598 lePheThrAspLysGlnGlnGluSerValMetGluIleThrSerAspVal 614
2263 GACATCAGAGGCAATGTCAGC.....CTTGC 2288
615 AspThrArgSerAspAlaAlaGlyHisGlyArgAspIleGluMetArgAl 631
2289 CGATCACGCTCATTTAAATCTCACAGGACTTGCCACACTCAACGGCAATC 2338
631 aAspGlyGluValAlaValAspAlaGlyValAspThrGlnTrpGlyAlaL 648
2339 TTAGTGCAGGCGGA.....GACACGCACTATACGGTTACG 2373
648 euMetAlaAspSerSerGlyGlnHisGlnAspGluGlySerThrLeuThr 664
2374 CGCAACGCCACCCAAAACGGCAACCTCAGCCTCGTGGGCAATGCCCAAGC 2423
665 LysThrGlyAlaGlyThrLeuGluLeuThrAlaSerGlyThrThrGlnSe 681
2424 AACA.....TTTAATCAAGCCACATTAAACGGCAACACATCG..... 2460
681 rAlaValArgValGluGluGlyThrLeuLysGlyAspValAlaAspIleL 698



```
      ::            ||||||:::  |||:::        |||
1133 alIleTyrGlnGlyValGlnGlnAspAspPheThrAlaAlaAsnArgAla 1149

4090 ....................GTCAATATCGCCACCCCGGGCCTTGCATT 4118
                         :::|||  |||
1150 ArgValSerGlnSerGlnGlyAspAspIleGlnThr.............. 1161

4119 CAACCGCTACCGCGCGGGCATTAAGGCAGATTATTCATTCAAACCGGCGC 4168
           |||  |||:::  :::::::::      :::  |||
1162 ..........ArgLeuGlyLeuHisSerGluTrp......ArgThrAlaV 1173

4169 AACACATTTCCATCACGCCTTATTTGAGCCTGTCC...TATACCGATGCC 4215
       |||     :::  |||  |||:::|||:::  |||  |||
1173 alHis......ValIleProThrLeuAspLeuAsnTyrTyrHisAspPro 1187

4216 GCTTCCGGCAAAGTCCGAACGCGCGTCAATACCGCCGTATTGGCGCAGGA 4265
       |||  :::::::::         :::|||     ::::::::::||
1188 HisSerThrGluIleGluGluAspGlySerThr......IleSerAspAs 1202

4266 TTTCGGCAAAACCCGC...AGTGCGGAATGGGGCGTAAACGCCGAAATCA 4312
     |    |||  |||  :::  :::  ||||||::::::::::|||:
1202 pAlaValLysGlnArgGlyGluIleLysValGlyValThrGlyAsnIleS 1219

4313 AAGGTTTCACGCTGTCCCTCCACGCTGCCGCCGCC......AAGGGGCCG 4356
     ::      :::||||||:::::::::::|||     ||||||.
1219 erGln...ArgValSerLeuArgGlySerValAlaTrpGlnLysGlySer 1234

4357 CAATTGGAAGCGCAGCACAGCGCGGGCATCAAATTAGGCTACCGCTGG 4404
     :::      ||||||  ::::::  :::::::::      :::|||
1235 AspAspPheAlaGlnThrAlaGlyPheLeuSerMetThrValLysTrp 1250

seq_name: pir2:D91018

seq_documentation_block:
hypothetical protein ECs3116 [imported] - Escherichia coli (strain O157:H7, substrain RI
C;Species: Escherichia coli
C;Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 18-Jul-2001
C;Accession: D91018
R;Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.
gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.
DNA Res. 8, 11-22, 2001
A;Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and genc
A;Reference number: A99629; MUID:21156231; PMID:11258796
A;Accession: D91018
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1250 <HAY>
A;Cross-references: GB:BA000007; PIDN:BAB36539.1; PID:g13362585; GSPDB:GN00154
A;Experimental source: strain O157:H7, substrain RIMD 0509952
C;Genetics:
A;Gene: ECs3116


alignment_scores:
            Quality:  290.50            Length:   1484
              Ratio:  0.441              Gaps:     73
 Percent Similarity:  44.340   Percent Identity: 19.946

alignment_block:
 US-09-303-518D-653 x D91018   ..

 Align seg 1/1  to: D91018  from: 1  to: 1250

    637 AACCGCGAAAGTTCATATCATATTGCAAGCGCATATTCT........... 675
        :::|||:::|||        |||:::::::::::::|||
     46 SerArgGlnSerLeuSerGlyIleThrGlnValTrpSerIleAlaAspGl 62

    676 ....TGGCTCGTCGGTGGCAATACCTTTGCACAAAATGGATCAGGTGGTG 721
            |||||||||  ::::::  :::::::::|||:::||||||||:
     62 yGlnTrpLeuValPheSerAsp...MetThrAsnAsnAlaSerGlyGlyA 78

    722 GCACAGTCAACTTAGGTAGCGAAAAAATTAAACATAGCCCATATGGTTTT 771
        :::::                                       |||
```

```
     78 laVal.....................................Phe 80

    772 TTACCAACAGGAGGCTCATTT..............GGCGACAGTGGCTC 806
        |||      |||:::::::|||              :::::::::|||
     81 LeuGlnGlnGlyAlaGluPheThrLeuSerProGluAsnGluThrGlyMe 97

    807 ACCAATGTTTATCTATGATGCCCAAAAGCAAAAGTGGTTAATTAATGGGG 856
        :::|||      ::::::                  ::::::|||
     97 tThrLeuPheAlaAsnAsnThr.................ValSerGlyG 108

    857 TATTGCAAACAGGCAACCCCTATATAGGAAAAAGCAATGGCTTCCAGCTA 906
            ::::::|||:::      :::::::|||:::|||:::
    108 luTyrAsnAsnGlyGlyAlaIlePheAlaLysGluAsnSerThrLeuAsn 124

    907 GTTCGTAAAGATTGGTTCTATGATGAAATCTTTGCTGGAGATACCCATTC 956
        :::                  |||   |||||||:::|||::::::   ::
    125 LeuThr..............AspValIlePheSerGlyAsnValAlaGl 136

    957 AGTATTCTACGAACCACATCAAAATGGGAAATACTTTTTTAACGACAATA 1006
        :   :::                  |||      ::::::::::::::::
    136 yGlyTyr.................GlyGlyAlaIleTyrSerSerGlyT 147

   1007 ATAATGGCGCAGGAAAAATCGATGCCAAACATAAACACTATTCTCTACCT 1056
        ::|||   :::|||   ||||||
    147 hrAsnAspThrGlyAlaIleAsp.......................... 154

   1057 TATAGATTAAAAACACGAACCGTTCAATTGTTTAATGTTTCTTTATCCGA 1106
            |||::::::  ::::::   :::   |||          :::::
    155 ......LeuArgValThrAsnAlaValPheArgAsnAsnIleAlaAsnAs 169

   1107 GACAGCAAGAGAACCTGTTTATCATGCTGCAGGTGGGGTCAACAGTTATC 1156
        :                ::|||         :::   :::
    169 pGlyLysGlyGlyAlaIleTyrThrIleAsnAsnAspIleTyrLeuSerA 186

   1157 GACCCAGACTGAATAATGGAGAA.........AATATTTCCTTTATTGAC 1197
              :::||||||::::::         :::   |||:::   |||
    186 spAspValPheAsnAsnAsnGlnAlaTyrThrSerThrSerTyrSerAsp 202

   1198 AAAGGAAAAGGTGAATTGATACTTACCAGCAAC.................. 1230
              |||   :::   :::|||:::|||
    203 GlyAspGlyGlyAlaIleAspValThrAspAsnAsnSerAspSerLysHi 219

   1231 ..................ATCAACCAAGGCGCGGGCGGTTTGTATTTTG 1261
                          ||||||:::   |||
    219 sProSerGlyTyrThrIleIleAsnAsnThrAla.............. 230

   1262 AGGGTAATTTTACGGTCTCGCCTAAAAACAACGAAACGTGGCAAGGCGCG 1311
              ||||||               |||:::|||   :::|||
    231 ........PheThr..............AsnAsnThrAlaGluGlyTyr 239

   1312 GGCGTTCATATC...AGTGATGGCAGTACCGTTACTTGGAAAGTAAACGG 1358
        |||      |||   ::::::::::::|||:::   :::   ::::::
    240 GlyGlyAlaIleTyrThrAsnSerAlaThrAlaProTyrLeuIleAspIl 256

   1359 CGTGGCAAACGACCGCCTGTCCAAAATCGGCAAAGGCACGCTGCTGGTTC 1408
           ::::::|||     |||:::  |||  |||:::|||:::   :
    256 eSerValAspAspSerTyrSerGlnAsnGly...GlyValLeuValAspG 272

   1409 AA.........GCCAAAGGGGAAAACCAAGGCTCGGTCAGC......... 1440
        ::         |||  |||  :::::::|||      |||
    272 luAsnAsnSerAlaAlaGlyTyrGlyAspGlyProSerSerAlaAlaGly 288

   1441 .....................................GTGGGCGACGG 1451
                                            ::::::|||||
    289 GlyPheMetTyrLeuGlyLeuSerGluValThrPheAspIleAlaAspGl 305

   1452 TAAAGTCATCTTAGATCAGCAGGCGGACGATCAAGGCAAAAAACAAGCCT 1501
        ||||::::::::      ::::::::::::::|||      :::::::
    305 yLysThrLeuValIleGlyAsnThrGluAsnAspGlyAlaValAspSerI 322

   1502 TTAGTGAAATCGGCTTGGTC......AGCGGCAGGGGGACGGTGCAACTG 1545
        :::::      ||||||:::      :::|||  |||  :::  |||
    322 leAlaGlyThrGlyLeuIleThrLysThrGlySerGlyAspLeuValLeu 338
```

```
1546 AATGCCGATAATCAGTTCAACCCCGACAAACTCTATTTCGGCTTTCGCGG 1595
     ||||||||||||
 339 AsnAlaAspAsn..................................... 342

1596 CGGACGTTTGGATTTGAACGGGCATTCGCTTTCGTTCCACCGCATTCAAA 1645
                                                      |
 343 .....................................................A 343

1646 ATACCGATGAAGGGGCGATGATTGTCAACCACAATCAAGACAAAGAATCC 1695
     ||         |||   |||   ::::::::::::::::
 343 snAspPheThrGlyGluMetGlnIleGluAsnGlyGlu............ 355

1696 ACCGTTACCATTACAGGCAATAAAGATATTACTACAACCGGCAATAACAA 1745
        ||||||:::      :::::::::::   :::::::|||:::::::
 356 ...ValThrLeuGlyArgSerAsnSerLeuMetAsnValGlyAspThrHi 371

1746 CAACTTGGATAGCAAAAAAGAAATTGCCTACAACGGTTGGTTTGGCGAGA 1795
     :      |||:::   ::::::   :::|||:::      :::|||:::
 371 sCysGlnAspAspProGlnAsp...CysTyrGlyLeuThrIleGlySerI 387

1796 AAGATGCAACCAAAACGAACGGGCGGCTCAATCTGAATTACCAACCGGAA 1845
       |||      :::::::::::::::|||||::::::          :::
 387 leAspLysTyrGlnAsnGlnAlaGluLeuAsnValGlySerThrGlnGln 403

1846 GAAGCGGATCGCACTTTA......CTGCTTTCCGGCGGAACAAATTTAAA 1889
                ::::::|||      :::   :::|||      |||:::::
 404 ThrPheAlaHisSerLeuThrGlyPheGlnAsnGlyThrLeuAsnIleAs 420

1890 C......GGCAATATCACGCAAACAAACGGCAAACTGTTTTTCAGCGGCA 1933
     :      ||||||:::|||   ::::::|||:::      |||:::|||
 420 pAlaGlyGlyAsnValThrValAsnGlnGlySer......PheAlaGly. 434

1934 GACCGACACCGCACGCCTACAATCATTTAGGAAGCGGGTGGTCAAAAATG 1983
                                                    :::
 435 .....................................ThrIle 436

1984 GAAGGTATCCCACAAGGAGAAATCGTGTGGGACAACGATTGGATCGACCG 2033
     ||||||      |||      |||:::   :::::::::::::::
 437 GluGlyAlaGlyGlnLeuThrIleAlaGlnAsnGlySerTyrVal..... 451

2034 CACATTTAAAGCGGAAAACTTCCATATTCAGGGCGGACAAGCGGTGGTTT 2083
               |||:::::::::   :::   |||      :::|||   :
 452 .LeuAlaGlyAlaGlnSerMetAlaLeuThrGlyAspIleValValAspA 468

2084 CCCGCAATGTTGCCAAAGTGGAAGGCGAT.....TGGCATTTA 2121
     ::      |||   ::::::|||||||||                |||
 468 laGlyAlaValLeuSerLeuGluGlyAspAlaAlaAspLeuAlaAlaLeu 484

2122 AGCAATCACGCCCAAGCAGTTTTCGGTGTCGCACCGCATCAAAGCCACAC 2171
     ::::::      |||:::::::      |||
 485 GlnAspAspProGlnSerIle......ValLeuAsnGlyGlyMetLeuAs 499

2172 AATCTGTACACGTTCGGACTGGACGGGTCTGACAAGTTGT.........A 2212
      :::         |||   |||   :::   ||||||
 499 pLeuSerAspPheSerThrTrpGlnSerGlyThrSerTyrLysAspGlyL 516

2213 CCGAAAAAACCATTACCGACGATAAAGTGATTGCTTCATTGAGCAAGACC 2262
       |||   :::   :::::      ||||||:::|||   :::   :::
 516 euGluValSerGlySerSerGlyThrValIleGlySerGlnAspValVal 532

2263 GACATCAGA..................................... 2271
     |||:::
 533 AspLeuAlaGlyGlyAsnAspMetHisIleGlyGlyAspGlyLysAspGl 549

2272 ..........................GGCAATGTCAGCCTTGCC...G 2290
                                |||:::|||||||||||||  |
 549 yValTyrValValIleAspAlaGlyAspGlyGlnValSerLeuAlaAsnA 566

2291 ATCACGCTCATTTAAATCTCACAGGACTTGCCACACTCAACGGCAATCTT 2340
     ||:::   :::|||:::   |||   :::|||      :::|||:::|||
 566 spAsnGlnTyrLeuGlyThrThrGlnIleAla......SerGlyThrLeu 580
```

```
2341 ..............AGTGCAGGCGGAGACACGCACTAT.......... 2364
                    |||      |||   |||||||||
 581 MetValSerAspAsnSerGlnLeuGlyTyrThrHisTyrAsnArgGlnVa 597

2365 .....................................................A 2365
                                                         :
 597 lIlePheThrAspLysProGlnGluSerValMetGluIleThrAlaAsnV 614

2366 CGGTTACGCGCAACGCCACC.......................... 2385
     ::   ||||||:::::::|||
 614 alAspThrArgSerThrThrThrGluHisGlyArgAspIleGluMetArg 630

2386 CAAAACGGCAACCTCAGCCTCGTGGGCAATGCCCAAGCAACATTTAATCA 2435
        :::|||:::::::::::   :::::::::::::::   ::::::
 631 AlaAspGlyGluValAlaValAspAlaGlyValAspThrGlnTrpGlyAl 647

2436 AGCCACATTAAACGGCAACACATCGGCTTCGGACAATGCTTCATTTAATC 2485
               :::::::::   :::   :::|||:::::::|||   :::|
 647 aLeuMetAlaAspSerSerGlyGlnHisGlnAspGluGlySer...ThrL 663

2486 TAAGCAACAACGCCGTACAAAACGGCAGTCTGACGCTTTCCGACAACGCT 2535
     ||:::::::::::::::   :::   :::||||||   |||   ::::::
 663 euThrLysThrGlyAlaGlyThrLeuGluLeuThrAlaSerGlyThrThr 679

2536 AAGGCA......AACGTAAGCCATTCCGCACTCAACGGCAATGTCTCCCT 2579
     ::::::      :::|||:::::::::::::|||:::|||:::|||
 680 GlnSerAlaValArgValGluGluGlyThrLeuGlnGlyAspVal..... 694

2580 AGCCGATAAGGCAGTATTCCATTTTGAAAACAGCCGCTTTACCGGAAAA. 2628
      ||||||      :::|||   :::   :::|||   ::::::|||
 695 .AlaAsp......IlePheProTyrAlaSerSerLeuTrpValGlyAspG 709

2629 ...........ATCAGCGGC........................... 2637
                 ::::::|||
 709 lyAlaThrPheValThrGlyAlaAspGlnAspIleGlnSerIleAspAla 725

2637 ..................................................... 2637

 726 ThrSerSerGlyThrIleAspIleSerAspGlyThrValLeuArgLeuTh 742

2638 .GGCAAGGATACGGCATTACACTTAAAAGACAGC............GAAT 2674
      |||:::||||||:::::::   |||:::   |||            :::
 742 rGlyGlnAspThrSerValAlaLeuAsnAlaSerLeuPheAsnCysAspG 759

2675 GGACGCTGCCGTCGGGCACGGAA............TTAGGCAATTTAAAC 2712
       ||||||   ::::::|||:::                |||:::||||||
 759 lyThrLeuValAsnAlaThrAspGlyValThrLeuThrGlyGluLeuAsn 775

2713 ..........................CTTGACAACGCCACCATTAC 2732
                                |||:::|||:::|||:::::
 776 ThrAsnLeuGluThrAspSerLeuThrTyrLeuSerAsnValThrValAs 792

2733 ACTCAATTCCGCCTATCGACACGATGCGGCAGGCGCGCAAACCGGCAGTG 2782
     :   |||   :::             |||::::::   |||:::|||   |
 792 nGlyAsnLeuThrAsnThrSerGlyAlaValSerLeuGlnAsnGlyValA 809

2783 CGGCAGATGCGCCGCGCCGCCGTTCGCGCCGTTCCCTATTATCCGTTACG 2832
     ||:::|||
 809 laGlyAsp..................................... 811

2833 CCGCCAACTTCGGCAGAATCCCGTTTCAACACGCTGACGGTAAACGGCAA 2882
                                |||||||||||||||||
 812 ..........................ThrLeuThrValAsnGlyAs 818

2883 ATTGAACGGTCAGGGAACATTCCGCTTTATGTCGGAACTCTTCGGC.... 2928
        :::|||   ||||||:::   :::   |||||||||   |||
 818 pTyrThrGlyGlyGlyThrLeuLeuLeuAspSerGluLeuAsnGlyAspA 835

2929 ..TACCGCAGCGGCAAATTGAAGCTGGCGGAAAGTTCCGAAGGCACTTAC 2976
           |||   :::|||   :::      ::::::   |||:::
 835 spSerValSerAspGlnLeuValMetAsnGlyAsnThrAlaGlyAsnThr 851

2977 ACCTTGGCTGTCAACAAT...ACCGGC...AACGAACCCGTAAGTCTCGA 3020
```

```
     |||:::::::|||||||:::   ||||||   :::||||||:::|||
 852 ThrValValValAsnSerIleThrGlyIleGlyGluProThrSerThrGl 868

3021 GCAATTGACGGTAGTGGAAGGAAAAGACAACACACCGCTGTCCGAAAATC 3070
          :::|||      :::   |||   |||   :::::::::|||
 868 yIleLysValValAspPheAlaAlaAspProThrGlnPheGlnAsnAsnA 885

3071 TTAATTTCACCCTGCAAAACGAA...CACGTCGATGCCGGCGCATGGCGT 3117
       :::|||:::|||   ::::::   :::|||:::   ||||||:::
 885 laGlnPheSerLeuAlaGlySerGlyTyrValAsnMetGlyAlaTyrAsp 901

3118 TATCAGCTTATCCGCAAAGACGGCGAGTTCCGCCTGCATAATCCGGTCAA 3167
     |||   |||::::::   ::::::::::::   |||
 902 TyrThrLeuValGluAspAsnAsnAspTrpTyrLeu..............  913

3168 AGAACAAGAGCTTTCCGACAAACTCGGCAAGGCGGGAGAAACAGAGGCCG 3217

 913 .................................................  913

3218 CCTTGACGGCAAAACAGGCACAACTTGCCGCCAAACAACAGGCGGAAAAA 3267

 913 .................................................  913

3268 GACAACGCGCAAAGCCTTGACGCGCTGATTGCGGCCGGGCGCAATGCCAC 3317

 913 .................................................  913

3318 CGAAAAGGCAGAAAGTGTTGCCGAACCGGCCCGGCAGGCAGGCGGGGAAA 3367
          :::::::::::::|||:::   |||
 914 ....ArgSerGlnGluValThrProPro.....................  921

3368 ATGCCGGCATTATGCAGGCGGAGGAAGAGAAAAAACGGGTGCAGGCGGAT 3417

 921 .................................................  921

3418 AAAGACACCGCCTTGGCGAAACAGCGCGAAGCGGAAACCCGGCCGGCTAC 3467
                                  :::          |||
 922 ............................SerProProAspProAspPr 928

3468 CACCGCCTTCCCCCGCGCCCGCCGCGCCCGCCGGGATTTGCCGCAACCGC 3517
     |||      |||        :::               |||:::|||:
 928 oThrProAspProAspProThrGlnAspProAspProThrProAspProG 945

3518 AGCCCCAACCGCAACCCCAACCGCAGCGCGACCTGATCAGCCGTTATGCC 3567
     ::|||   |||      ||||||      :::
 945 luProThrProAlaTyrGlnProValLeuAsn..................  955

3568 AATAGCGGTTTGAGTGAATTTTCCGCCACGCTCAACAGCGTTTTCGCCGT 3617
               ::::::   ::::::   ||||||::::::   |||::
 956 ............AlaLysValGlyGlyTyrLeuAsnAsnLeuArgAlaAl 968

3618 ACAGGACGAATTGGACCGCGTGTTTGCCGAAGACCGCCGCAACGCCGTTT 3667
     :              :::::::::|||      :::|||||||:::   :::
 968 a............AsnGlnAlaPheMetMetGluArgArgAspHisAlaG 981

3668 GGACAAGCGGC.....................ATCCGGGACACCAAACAC 3696
        :::|||                       |||            |||
 981 lyGlyAspGlyGlnThrLeuAsnLeuArgValIleGlyGlyAspTyrHis 997

3697 TACCGTTCGCAAGATTTCCGCGCCTACCGCCAACAAACCGACCTGCGCCA 3746
     |||   :::           |||      |||:::   |||
 998 TyrThrAlaAlaGlyGlnLeuAla......GlnHisGluAspThrSerTh 1012

3747 AATCGGTATGCAGAAAAACCTCGGCAGCGGGCGC.............. 3780
      :::   ::::::   :::|||   |||||||||
1012 rValGlnLeuSerGlyAspLeuPheSerGlyArgTrpGlyThrAspGlyG 1029

3781 ........GTCGGCATCCTG......TTTTCGCACAACCGGACCGGAAAC 3816
           :::||||||:::      :::|||   |||:::   |||:::
1029 luTrpMetLeuGlyIleValGlyGlyTyrSerAspAsnGln...GlyAsp 1044

3817 ACCTTCGACGACGGCATCGGCAACTCGGCA...CGGCTTGCCCACGGTGC 3863
     :::   ::::::      |||:::   |||   :::      ||||||
```

```
1045 SerArgSerSerMetThrGlyThrArgAlaAspAsnGlnAsnHisGly.. 1060

3864 CGTTTTCGGGCAATACGGCATCGGCAGGTTCGACATCGGCATCAGCGCGG 3913
                  |||::::::|||      :::      ::::::   |
1061 .............TyrAlaValGlyLeuThrSerSerTrpPheGlnHisG 1073

3914 GC.....................................GCGGGT 3921
     ||                                     |||
1073 lyLysGlnLysGlnGlyAlaTrpLeuAspAsnTrpLeuGlnTyrAlaTrp 1089

3922 TTTAGTAGC......GGCAGCCTTTCAGACGGCATCAGAGGCAAAATCCG 3965
     |||||||::      ::::::   :::|||||||:::
1090 PheSerAsnAspValSerGluHisGluAspGlyValAsp........... 1102

3966 CCGCCGCGTGCTGCATTAC.........GGCATTCAGGCAAGATACCGCG 4006
                  ||||||         ||||||   |||      :::|
1103 .............HisTyrHisSerSerGlyIleIleAlaSerLeuGluA 1115

4007 CAGGTTTC...............GGCGGATTCGGCATCGAACCGCACATC 4041
     |||||:::                |||      ||||||||||:::
1115 laGlyTyrGlnTrpLeuProGlyArgGlyValValIleGluProGlnAla 1131

4042 GGCGCAACGCGCTATTTCGTCCAAAAAGCGGATTACCGATACGAAAAC.. 4089
        :::           |||||||:::   |||:::          |||
1132 GlnValIleTyrGlnGlyValGlnGlnAspAspPheThrAlaAlaAsnAr 1148

4090 .........................GTCAATATCGCCACCCCGGGCCTTG 4114
                                     :::|||   |||
1148 gAlaArgValSerGlnSerGlnGlyAspAspIleGlnThr.......... 1161

4115 CATTCAACCGCTACCGCGCGGGCATTAAGGCAGATTATTCATTCAAACCG 4164
                  |||   |||:::   :::::::::      :::
1162 .............ArgLeuGlyLeuHisSerGluTrp......ArgThr 1171

4165 GCGCAACACATTTCCATCACGCCTTATTTGAGCCTGTCC...TATACCGA 4211
     |||   |||      :::   |||   |||:::|||:::   |||   ||
1172 AlaValHis......ValIleProThrLeuAspLeuAsnTyrTyrHisAs 1186

4212 TGCCGCTTCCGGCAAAGTCCGAACGCGCGTCAATACCGCCGTATTGGCGC 4261
     |      |||   :::::::::          :::::::|||      :::::::
1186 pProHisSerThrGluIleGluGluAspAlaSerThr......IleSerA 1201

4262 AGGATTTCGGCAAAACCCGC...AGTGCGGAATGGGGCGTAAACGCCGAA 4308
     ::|||      |||   |||   :::   :::   ||||||::::::::
1201 spAspAlaValLysGlnArgGlyGluIleLysValGlyValThrGlyAsn 1217

4309 ATCAAAGGTTTCACGCTGTCCCTCCACGCTGCCGCCGCC......AAGGG 4352
     |||:::         :::||||||:::::::::::::|||      |||||
1218 IleSerGln...ArgValSerLeuArgGlySerValAlaTrpGlnLysGl 1233

4353 GCCGCAATTGGAAGCGCAGCACAGCGCGGGCATCAAATTAGGCTACCGCT 4402
     |   :::      ||||||   ::::::   :::::::::      :::|
1233 ySerAspAspPheAlaGlnThrAlaGlyPheLeuSerMetThrValLysT 1250

4403 GG 4404
     ||
1250 rp 1250
```

seq_name: pir2:F85862

seq_documentation_block:
hypothetical protein yfaL [imported] - Escherichia coli (strain O157:H7, substrain ED
C;Species: Escherichia coli
C;Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 14-Sep-2001
C;Accession: F85862
R;Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; May
iller, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apoda
Nature 409, 529-533, 2001
A;Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
A;Reference number: A85480; MUID:21074935; PMID:11206551
A;Accession: F85862
A;Status: preliminary
A;Molecule type: DNA

```
A;Residues: 1-1250 <STO>
A;Cross-references: GB:AE005174; NID:g12516568; PIDN:AAG57362.1; GSPDB:GN00145; UWGP:Z34
A;Experimental source: strain O157:H7, substrain EDL933
C;Genetics:
A;Gene: yfaL


alignment_scores:
            Quality:  290.50            Length:   1484
              Ratio:  0.441               Gaps:     73
 Percent Similarity:  44.340   Percent Identity: 19.946

alignment_block:
 US-09-303-518D-653 x F85862   ..

 Align seg 1/1  to: F85862  from: 1  to: 1250

     637 AACCGCGAAAGTTCATATCATATTGCAAGCGCATATTCT..... 675
         ::::|||:::|||         |||:::::::::::|||
      46 SerArgGlnSerLeuSerGlyIleThrGlnValTrpSerIleAlaAspGl 62

     676 ....TGGCTCGTCGGTGGCAATACCTTTGCACAAAATGGATCAGGTGGTG 721
             |||||||||   ::::::   ::::::::::|||:::|||||||||:
      62 yGlnTrpLeuValPheSerAsp...MetThrAsnAsnAlaSerGlyGlyA 78

     722 GCACAGTCAACTTAGGTAGCGAAAAAATTAAACATAGCCCATATGGTTTT 771
         :::::                                          |||
      78 laVal.........................................Phe 80

     772 TTACCAACAGGAGGCTCATTT..............GGCGACAGTGGCTC 806
         |||      |||:::::::|||               :::::::::|||
      81 LeuGlnGlnGlyAlaGluPheThrLeuSerProGluAsnGluThrGlyMe 97

     807 ACCAATGTTTATCTATGATGCCCAAAAGCAAAAGTGGTTAATTAATGGGG 856
             :::|||      ::::::                 ::::::|||
      97 tThrLeuPheAlaAsnAsnThr.................ValSerGlyG 108

     857 TATTGCAAACAGGCAACCCCTATATAGGAAAAAGCAATGGCTTCCAGCTA 906
             ::::::|||:::      :::::::|||:::|||:::
     108 luTyrAsnAsnGlyGlyAlaIlePheAlaLysGluAsnSerThrLeuAsn 124

     907 GTTCGTAAAGATTGGTTCTATGATGAAATCTTTGCTGGAGATACCCATTC 956
         :::                  |||   ||||||:::|||::::::   ::
     125 LeuThr..............AspValIlePheSerGlyAsnValAlaGl 136

     957 AGTATTCTACGAACCACATCAAAATGGGAAATACTTTTTTAACGACAATA 1006
         :   :::                  |||      ::::::::::::::::
     136 yGlyTyr.................GlyGlyAlaIleTyrSerSerGlyT 147

    1007 ATAATGGCGCAGGAAAAATCGATGCCAAACATAAACACTATTCTCTACCT 1056
         ::|||   :::|||   ||||||
     147 hrAsnAspThrGlyAlaIleAsp........................... 154

    1057 TATAGATTAAAAACACGAACCGTTCAATTGTTTAATGTTTCTTTATCCGA 1106
             |||::::::   ::::::   :::   |||        :::::
     155 ......LeuArgValThrAsnAlaValPheArgAsnAsnIleAlaAsnAs 169

    1107 GACAGCAAGAGAACCTGTTTATCATGCTGCAGGTGGGGTCAACAGTTATC 1156
         :                :::|||         :::   :::
     169 pGlyLysGlyGlyAlaIleTyrThrIleAsnAsnAspIleTyrLeuSerA 186

    1157 GACCCAGACTGAATAATGGAGAA.........AATATTTCCTTTATTGAC 1197
                 :::||||||::::::         :::   |||:::   |||
     186 spAspValPheAsnAsnAsnGlnAlaTyrThrSerThrSerTyrSerAsp 202

    1198 AAAGGAAAAGGTGAATTGATACTTACCAGCAAC.................. 1230
                   |||   :::   :::|||:::|||
     203 GlyAspGlyGlyAlaIleAspValThrAspAsnAsnSerAspSerLysHi 219

    1231 ..................ATCAACCAAGGCGCGGGCGGTTTGTATTTTG 1261
                            |||||||:::   |||
     219 sProSerGlyTyrThrIleIleAsnAsnThrAla.............. 230

    1262 AGGGTAATTTTACGGTCTCGCCTAAAAACAACGAAACGTGGCAAGGCGCG 1311
                |||||                |||:::|||   :::|||
     231 ........PheThr..............AsnAsnThrAlaGluGlyTyr 239

    1312 GGCGTTCATATC...AGTGATGGCAGTACCGTTACTTGGAAAGTAAACGG 1358
         |||      |||   ::::::::::::|||:::   :::   ::::::
     240 GlyGlyAlaIleTyrThrAsnSerAlaThrAlaProTyrLeuIleAspIl 256

    1359 CGTGGCAAACGACCGCCTGTCCAAAATCGGCAAAGGCACGCTGCTGGTTC 1408
             ::::::|||      |||:::   |||   |||:::|||:::   :
     256 eSerValAspAspSerTyrSerGlnAsnGly...GlyValLeuValAspG 272

    1409 AA.........GCCAAAGGGGAAAACCAAGGCTCGGTCAGC......... 1440
         ::         |||   |||   ::::::|||      |||
     272 luAsnAsnSerAlaAlaGlyTyrGlyAspGlyProSerSerAlaAlaGly 288

    1441 .....................................GTGGGCGACGG 1451
                                                ::::::|||||
     289 GlyPheMetTyrLeuGlyLeuSerGluValThrPheAspIleAlaAspGl 305

    1452 TAAAGTCATCTTAGATCAGCAGGCGGACGATCAAGGCAAAAAACAAGCCT 1501
         ||||::::::::::       ::::::::::::::::|||      :::::::
     305 yLysThrLeuValIleGlyAsnThrGluAsnAspGlyAlaValAspSerI 322

    1502 TTAGTGAAATCGGCTTGGTC......AGCGGCAGGGGGACGGTGCAACTG 1545
         :::::      |||||||::      :::|||   |||   :::   |||
     322 leAlaGlyThrGlyLeuIleThrLysThrGlySerGlyAspLeuValLeu 338

    1546 AATGCCGATAATCAGTTCAACCCCGACAAACTCTATTTCGGCTTTCGCGG 1595
         ||||||||||||
     339 AsnAlaAspAsn..................................... 342

    1596 CGGACGTTTGGATTTGAACGGGCATTCGCTTTCGTTCCACCGCATTCAAA 1645
                                                         |
     343 .............................................A 343

    1646 ATACCGATGAAGGGGCGATGATTGTCAACCACAATCAAGACAAAGAATCC 1695
         ||         |||   |||   ::::::::::::::::
     343 snAspPheThrGlyGluMetGlnIleGluAsnGlyGlu............ 355

    1696 ACCGTTACCATTACAGGCAATAAAGATATTACTACAACCGGCAATAACAA 1745
             ||||||:::      :::::::::::: ' ::::::|||::::::::
     356 ...ValThrLeuGlyArgSerAsnSerLeuMetAsnValGlyAspThrHi 371

    1746 CAACTTGGATAGCAAAAAAGAAATTGCCTACAACGGTTGGTTTGGCGAGA 1795
         :      |||:::   ::::::   :::|||:::      :::|||:::
     371 sCysGlnAspAspProGlnAsp...CysTyrGlyLeuThrIleGlySerI 387

    1796 AAGATGCAACCAAAACGAACGGGCGGCTCAATCTGAATTACCAACCGGAA 1845
           |||      ::::::::::::::::::||||||::::::          :::
     387 leAspLysTyrGlnAsnGlnAlaGluLeuAsnValGlySerThrGlnGln 403

    1846 GAAGCGGATCGCACTTTA......CTGCTTTCCGGCGGAACAAATTTAAA 1889
                 ::::::|||      :::   :::|||      |||:::::
     404 ThrPheAlaHisSerLeuThrGlyPheGlnAsnGlyThrLeuAsnIleAs 420

    1890 C......GGCAATATCACGCAAACAAACGGCAAACTGTTTTTCAGCGGCA 1933
         :      ||||||:::|||   ::::::|||:::      |||:::|||
     420 pAlaGlyGlyAsnValThrValAsnGlnGlySer......PheAlaGly. 434

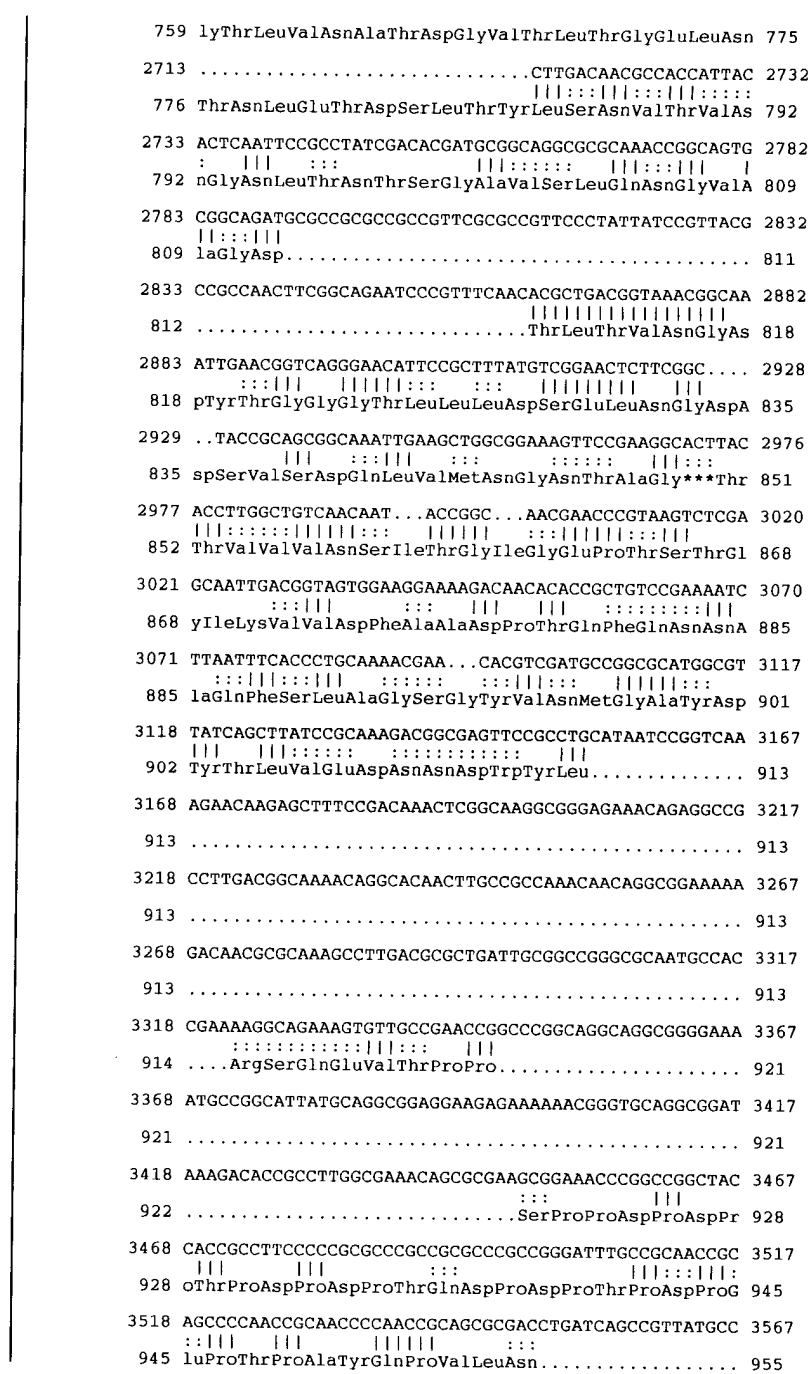
    1934 GACCGACACCGCACGCCTACAATCATTTAGGAAGCGGGTGGTCAAAAATG 1983
                                                        :::
     435 .........................................ThrIle 436

    1984 GAAGGTATCCCACAAGGAGAAATCGTGTGGGACAACGATTGGATCGACCG 2033
         ||||||      |||      |||:::   ::::::::::::::
     437 GluGlyAlaGlyGlnLeuThrIleAlaGlnAsnGlySerTyrVal..... 451

    2034 CACATTTAAAGCGGAAAACTTCCATATTCAGGGCGGACAAGCGGTGGTTT 2083
                 |||::::::::::   :::   |||      :::|||   :
     452 .LeuAlaGlyAlaGlnSerMetAlaLeuThrGlyAspIleValValAspA 468

    2084 CCCGCAATGTTGCCAAAGTGGAAGGCGAT............TGGCATTTA 2121
```


laGlyAlaValLeuSerLeuGluGlyAspAlaAlaAspLeuAlaAlaLeu
GlnAspAspProGlnSerIle.....ValLeuAsnGlyGlyMetLeuAs
pLeuSerAspPheSerThrTrpGlnSerGlyThrSerTyrLysAspGlyL
euGluValSerGlySerSerGlyThrValIleGlySerGlnAspValVal
AspLeuAlaGlyGlyAsnAspMetHisIleGlyGlyAspGlyLysAspGl
yValTyrValValIleAspAlaGlyAspGlyGlnValSerLeuAlaAsnA
spAsnGlnTyrLeuGlyThrThrGlnIleAla.....SerGlyThrLeu
MetValSerAspAsnSerGlnLeuGlyTyrThrHisTyrAsnArgGlnVa
lIlePheThrAspLysProGlnGluSerValMetGluIleThrAlaAsnV
alAspThrArgSerThrThrThrGluHisGlyArgAspIleGluMetArg
AlaAspGlyGluValAlaValAspAlaGlyValAspThrGlnTrpGlyAl
aLeuMetAlaAspSerSerGlyGlnHisGlnAspGluGlySer...ThrL
euThrLysThrGlyAlaGlyThrLeuGluLeuThrAlaSerGlyThrThr
GlnSerAlaValArgValGluGluGlyThrLeuGlnGlyAspVal.....
.AlaAsp.....IlePheProTyrAlaSerSerLeuTrpValGlyAspG
lyAlaThrPheValThrGlyAlaAspGlnAspIleGlnSerIleAspAla
ThrSerSerGlyThrIleAspIleSerAspGlyThrValLeuArgLeuTh
rGlyGlnAspThrSerValAlaLeuAsnAlaSerLeuPheAsnCysAspG


lyThrLeuValAsnAlaThrAspGlyValThrLeuThrGlyGluLeuAsn
ThrAsnLeuGluThrAspSerLeuThrTyrLeuSerAsnValThrValAs
nGlyAsnLeuThrAsnThrSerGlyAlaValSerLeuGlnAsnGlyValA
laGlyAsp
ThrLeuThrValAsnGlyAs
pTyrThrGlyGlyGlyThrLeuLeuLeuAspSerGluLeuAsnGlyAspA
spSerValSerAspGlnLeuValMetAsnGlyAsnThrAlaGly***Thr
ThrValValValAsnSerIleThrGlyIleGlyGluProThrSerThrGl
yIleLysValValAspPheAlaAlaAspProThrGlnPheGlnAsnAsnA
laGlnPheSerLeuAlaGlySerGlyTyrValAsnMetGlyAlaTyrAsp
TyrThrLeuValGluAspAsnAsnAspTrpTyrLeu
ArgSerGlnGluValThrProPro
oThrProAspProAspProThrGlnAspProAspProThrProAspProG
luProThrProAlaTyrGlnProValLeuAsn

```
3568 AATAGCGGTTTGAGTGAATTTTCCGCCACGCTCAACAGCGTTTTCGCCGT 3617
                :::::   ::::::   ||||||:::::   ||||:
 956 ...........AlaLysValGlyGlyTyrLeuAsnAsnLeuArgAlaAl 968

3618 ACAGGACGAATTGGACCGCGTGTTTGCCGAAGACCGCCGCAACGCCGTTT 3667
     :           :::::::::|||      :::|||||||:::  :::
 968 a...........AsnGlnAlaPheMetMetGluArgArgAspHisAlaG 981

3668 GGACAAGCGGC.....................ATCCGGGACACCAAACAC 3696
       :::|||                       |||           |||
 981 lyGlyAspGlyGlnThrLeuAsnLeuArgValIleGlyGlyAspTyrHis 997

3697 TACCGTTCGCAAGATTTCCGCGCCTACCGCCAACAAACCGACCTGCGCCA 3746
     |||   :::          |||      |||:::  |||
 998 TyrThrAlaAlaGlyGlnLeuAla......GlnHisGluAspThrSerTh 1012

3747 AATCGGTATGCAGAAAAACCTCGGCAGCGGGCGC.............. 3780
      :::   ::::::   :::|||   |||||||||
1012 rValGlnLeuSerGlyAspLeuPheSerGlyArgTrpGlyThrAspGlyG 1029

3781 ........GTCGGCATCCTG......TTTTCGCACAACCGGACCGGAAAC 3816
           :::||||||:::      :::|||   |||:::   |||:::
1029 luTrpMetLeuGlyIleValGlyGlyTyrSerAspAsnGln...GlyAsp 1044

3817 ACCTTCGACGACGGCATCGGCAACTCGGCA...CGGCTTGCCCACGGTGC 3863
     :::   ::::::      |||:::   |||   :::      ||||||
1045 SerArgSerSerMetThrGlyThrArgAlaAspAsnGlnAsnHisGly.. 1060

3864 CGTTTTCGGGCAATACGGCATCGGCAGGTTCGACATCGGCATCAGCGCGG 3913
               |||::::::|||      :::      ::::::   |
1061 .............TyrAlaValGlyLeuThrSerSerTrpPheGlnHisG 1073

3914 GC.........................................GCGGGT 3921
     ||                                          |||
1073 lyLysGlnLysGlnGlyAlaTrpLeuAspAsnTrpLeuGlnTyrAlaTrp 1089

3922 TTTAGTAGC......GGCAGCCTTTCAGACGGCATCAGAGGCAAAATCCG 3965
     ||||||:::      ::::::   :::|||||||:::
1090 PheSerAsnAspValSerGluHisGluAspGlyValAsp.......... 1102

3966 CCGCCGCGTGCTGCATTAC.........GGCATTCAGGCAAGATACCGCG 4006
                |||||          ||||||   |||      :::|
1103 .............HisTyrHisSerSerGlyIleIleAlaSerLeuGluA 1115

4007 CAGGTTTC...............GGCGGATTCGGCATCGAACCGCACATC 4041
     |||||:::                |||      ||||||||||:::
1115 laGlyTyrGlnTrpLeuProGlyArgGlyValValIleGluProGlnAla 1131

4042 GGCGCAACGCGCTATTTCGTCCAAAAAGCGGATTACCGATACGAAAAC.. 4089
       :::            ||||||:::   |||:::          |||
1132 GlnValIleTyrGlnGlyValGlnGlnAspAspPheThrAlaAlaAsnAr 1148

4090 .........................GTCAATATCGCCACCCCGGGCCTTG 4114
                                ::::|||   |||
1148 gAlaArgValSerGlnSerGlnGlyAspAspIleGlnThr.......... 1161

4115 CATTCAACCGCTACCGCGCGGGCATTAAGGCAGATTATTCATTCAAACCG 4164
                  |||   |||:::   :::::::::      :::
1162 .............ArgLeuGlyLeuHisSerGluTrp......ArgThr 1171

4165 GCGCAACACATTTCCATCACGCCTTATTTGAGCCTGTCC...TATACCGA 4211
     |||   |||      :::   |||   |||:::|||:::   |||  ||
1172 AlaValHis......ValIleProThrLeuAspLeuAsnTyrTyrHisAs 1186

4212 TGCCGCTTCCGGCAAAGTCCGAACGCGCGTCAATACCGCCGTATTGGCGC 4261
     |      |||   :::::::::      ::::::|||      :::::::
1186 pProHisSerThrGluIleGluGluAspAlaSerThr......IleSerA 1201

4262 AGGATTTCGGCAAAACCCGC...AGTGCGGAATGGGGCGTAAACGCCGAA 4308
     ::|||      |||   |||   :::   :::   ||||||:::::::::
1201 spAspAlaValLysGlnArgGlyGluIleLysValGlyValThrGlyAsn 1217

4309 ATCAAAGGTTTCACGCTGTCCCTCCACGCTGCCGCCGCC......AAGGG 4352
     |||:::          :::||||||:::::::::::|||      |||||
1218 IleSerGln...ArgValSerLeuArgGlySerValAlaTrpGlnLysGl 1233

4353 GCCGCAATTGGAAGCGCAGCACAGCGCGGGCATCAAATTAGGCTACCGCT 4402
     |   :::      ||||||   ::::::   :::::::::      :::|
1233 ySerAspAspPheAlaGlnThrAlaGlyPheLeuSerMetThrValLysT 1250

4403 GG 4404
     ||
1250 rp 1250
```

seq_name: pir2:B43855

seq_documentation_block:
high-molecular-weight surface-exposed protein - Haemophilus influenzae
C;Species: Haemophilus influenzae
C;Date: 10-Mar-1993 #sequence_revision 18-Nov-1994 #text_change 18-Nov-1994
C;Accession: B43855
R;Barenkamp, S.J.; Leininger, E.
Infect. Immun. 60, 1302-1313, 1992
A;Title: Cloning, expression, and DNA sequence analysis of genes encoding nontypeable
detella pertussis.
A;Reference number: A43855; MUID:92192797
A;Accession: B43855
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1477 <BAR>
A;Note: sequence inconsistent with the nucleotide translation
A;Note: sequence extracted from NCBI backbone (NCBIN:89237, NCBIP:89240)

alignment_scores:
Quality: 279.50 Length: 1319
Ratio: 0.437 Gaps: 61
Percent Similarity: 48.446 Percent Identity: 20.015

alignment_block:
US-09-303-518D-653 x B43855 ..

Align seg 1/1 to: B43855 from: 1 to: 1477

```
 706 AATGGATCAGGTGGTGGCACAGTCAACTTAGGTAGCGAAAAAATTAAACA 755
     ||||||      ::::::|||:::::::::::::::|||::||||||
 165 AsnGlyPheThrAlaSerThrLeuAspIleSerAsnGluAsnIleLys.. 180

 756 TAGCCCATATGGTTTTTTACCAACAGGAGGCTCATTTGGCGACAGTGGCT 805
                                                      :
 181 ..................................................A 181

 806 CACCAATGTTTATCTATGATGCCCAAAAGCAAAAGTGGTTAATTAATGGG 855
     ::      |||   ::::::     |||:::|||   |||   :::
 181 laArgAsnPheThrPheGluGlnThrLysAspLysAlaLeuAlaGlu... 196

 856 GTATTGCAAACAGGCAACCCCTATATAGGAAAAAGCAATGGCTTCCAGCT 905
     :::::::::   |||         :::|||||||:::::::::   :::||
 197 IleValAsnHisGlyLeuIleThrValGlyLysAspGlySerValAsnLe 213

 906 AGTTCGTAAAGATTGGTTCTATGATGAAATCTTTGCTGGAGATACCCATT 955
     |:::
 213 uIle.............................................. 214

 956 CAGTATTCTACGAACCACATCAAAATGGGAAATACTTTTTTAACGACAAT 1005
                               :::||||||
 215 .......................GlyGlyLysVal............... 218

1006 AATAATGGCGCAGGAAAAATCGATGCCAAACATAAACACTATTCTCTACC 1055
        ::::::   |||   |||:::::::::            ||||||
 219 ...LysAsnGluGlyValIleSerValAsnGlyGlySerIleSerLeuLe 234

1056 TTATAGATTAAAAACACGAACCGTTCAATTGTTTAATGTTTCTTTATCCG 1105
               |||            :::::::::|||   :::::::::
 234 uAlaGlyGlnLysIleThrIleSerAspIleIleAsnProThrIleThrT 251
```



```
                                :::::::::   :::|||   |||:::
 787 spAspAlaPheLysIleAsnLysAspLeuThrIleAsnAlaThrAsnSer 803

2569 AATGTCTCCCTAGCCGATAAGGCAGTATTCCATTTTGAA......:.... 2607
     |||   ||||||   :::                  ::::::
 804 AsnPheSerLeuArgGlnThrLysAspAspPheTyrAspGlyTyrAlaAr 820

2608 .........AACAGCCGCTTTACCGGAAAAATCAGCGGCGGCAAGGATA 2647
             ||||||   ::::::   :::|||   ||||||:::   |
 820 gAsnAlaIleAsnSerThrTyrAsnIleSerIleLeuGlyGlyAsnValT 837

2648 CGGCATTACACTTAAAAGACAGCGAATGGACGCTGCCGTCGGGCACGGAA 2697
     ||   |||      ::::::|||               |||::::::
 837 hr...LeuGlyGlyGlnAsnSer.............SerSerSerIle 847

2698 TTAGGCAATTTAAACCTTGACAAC...GCCACCATTACACTCAATTCCGC 2744
        ||||||:::::::::::::::   |||:::::::||||||
 848 ThrGlyAsnIleThrIleGluLysAlaAlaAsnValThrLeu........ 861

2745 CTATCGACACGATGCGGCAGGCGCGCAAACCGGCAGTGCGGCAGATGCGC 2794
                                        :::|||   :::||||
 862 ................................GluAlaAsnAsnAlaP 867

2795 CGCGCCGCCGTTCGCGCCGTTCCCTATTATCCGTTACGCCGCCAACTTCG 2844
     ||::::::::::::   |||:::   :::
 867 roAsnGlnGlnAsnIleArgAspArgVal..................... 876

2845 GCAGAATCCCGTTTCAACACGCTGACGGTAAACGGCAAATTGAACGGTCA 2894
             :::::::::::|||   ||||||||||:::|||:::  :
 877 ......IleLysLeuGlySerLeuLeuValAsnGlySerLeuSerLeuTh 891

2895 GGGAACATTCCGCTTTATGTCGGAACTCTTCGGCTACCGCAGCGGCAAAT 2944
      |||           :::                        :::|||:::|
 891 rGlyGluAsnAlaAspIle.....................LysGlyAsnL 901

2945 TGAAGCTGGCGGAAAGTTCC.......................... 2964
     ||   ::::::||||||:::
 901 euThrIleSerGluSerAlaThrPheLysGlyLysThrArgAspThrLeu 917

2965 ......GAAGGCACTTACACCTTGGCTGTCAACAATACCGGCAACGAACC 3008
            |||:::::::|||           |||:::|||:::
 918 AsnIleThrGlyAsnPheThr............AsnAsnGlyThrAlaGl 930

3009 CGTAAGTCTCGAGCAATTGACGGTAGTGGAAGGAAAAGACAACACACCGC 3058
      :::::::::   |||   :::|||      |||      |||:::
 930 uIleAsnIleThrGlnGlyValValLysLeuGly......AsnValThrA 945

3059 TGTCCGAAAATCTTAATTTCACCCTGCAAAACGAACACGTCGATGCCGGC 3108
       :::   :::|||||||:::|||   :::   ::::::   :::   :::
 945 snAspGlyAspLeuAsnIleThrThrHisAlaLysArgAsnGlnArgSer 961

3109 GCATGGCGTTATCAGCTTATCCGCAAAGACGGCGAGTTCCGCCTGCATAA 3158
                :::::::|||:::|||   |||:::::::::::   ::
 962 IleIleGlyGlyAspIleIleAsnLysLysGlySerLeuAsnIleThrAs 978

3159 TCCGGTCAAAGAACAAGAGCTT.......................... 3180
     :      ::::::   |||:::                          .
 978 pSerAsnAsnAspAlaGluIleGlnIleGlyGlyAsnIleSerGlnLysG 995

3181 ..................TCCGACAAACTC................... 3192
                        ||||||||||:::
 995 luGlyAsnLeuThrIleSerSerAspLysIleAsnIleThrLysGlnIle 1011

3193 ............GGCAAGGCGGGAGAAACAGAGGCCGCCTTGACGGCAAA 3230
                |||      ||||||   :::::::   :::::::
1012 ThrIleLysLysGlyIleAspGlyGluAspSerSerSerAspAlaThrSe 1028

3231 ACAGGCACAACTTGCCGCCAAACAACAGGCGGAAAAAGACAACGCGCAAA 3280
     ::::|||:::|||:::   |||   :::      |||
1028 rAsnAlaAsnLeuThrIleLysThrLysGluLeuLys............L 1041

3281 GCCTTGACGCGCTGATTGCGGCCGGGCGCAATGCCACCGAAAAGGCAGAA 3330
          :::   |||      :::|||   |||         |||||||||
```

```
1041 euThrGluAspLeuSerIleSerGlyPheAsn.........LysAlaGlu 1054

3331 AGTGTTGCCGAACCGGCCCGGCAG......GCAGGCGGGGAAAATGCCGG 3374
        :::|||:::   :::|||:::         |||:::::::||   ||
1055 IleThrAlaLysAspGlyArgAspLeuThrIleGlyAsnSerAsnAspGl 1071

3375 CATTATGCAGGCGGAGGAAGAGAAAAAACGGGTGCAGGCGGATAAAGACA 3424
     |         ||||||   :::          :::      ||||||:
1071 yAsnSerGlyAlaGluAlaLysThrValThrPheAsnAsnValLysAspS 1088

3425 CCGCCTTGGCGAAACAG..............CGCGAAGCGGAAACCCGG 3459
     ::   ::::::   :::                  ::::::::::::::
1088 erLysIleSerAlaAspGlyHisAsnValThrLeuAsnSerLysValLys 1104

3460 CCGGCTACCACCGCCTTCCCCCGCGCCCGCCGCGCCCGCCGGGATTTGCC 3509
        :::::::::          |||      ::::::   :::|||
1105 ThrSerSerSerAsnGlyGlyArgGluSerAsnSerAspAsnAspThrGl 1121

3510 GCAACCGCAGCCCCAACCGCAACCCCAACCGCAGCGCGACCTGATCAGC. 3558
                           :::   :::   ::::::|||:::   |||
1121 yLeuThrIleThrAlaLysAsnValGluValAsnLysAspIleThrSerL 1138

3559 ..CGTTATGCCAATAGCGGTTTGAGTGAATTTTCCGCCACGCTCAACAGC 3606
       :::   :::|||         ||||||      :::||| ::::::::
1138 euLysThrValAsnIleThrAlaSerGluLysValThrThr.ThrAlaGl 1154

3607 GTTTTCGCCGTACAGGACGAATTGGACCGCGTGTTTGCCGAAGACCGCCG 3656
     :   :::          ||||||      |||   :::         ::::
1154 ySerThrIleAsnAlaThrAsnGlyLysAlaSerIleThrThrLysThrG 1171

3657 CAACGCCGTTTGGACAAGCGGCATCCGGGACACCAAACACTACCGTTCGC 3706
     ::         |||      :::::::::|||   ::::::::::::
1171 lyAspIleSerGlyThrIleSerGlyAsnThrValSerValSerAlaThr 1187

3707 ...AAGATTTCCGCGCCTACCGCCAACAAACCGACCTGCGCCAAATCGGT 3753
           :::::::::   ::::::::::|||      :::   ::::::
1188 ValAspLeuThrThrLysSerGlySerLysIleGluAlaLysSerGlyGl 1204

3754 ATGCAGAAAAACCTCGGCAGCGGGCGCGTCGGCATCCTGTTTTCGCACAA 3803
       ::::::   |||||||||:::|||:::   ::::::          |||
1204 uAlaAsnValThrSerAlaThrGlyThrIleGlyGly.........ThrI 1218

3804 CCGGACCGGAAACACCTTCGACGACGGCATCGGCA.............. 3838
       :::   :::|||   ::::::||||||:::|||
1218 leSerGlyAsnThrValAsnValThrAlaAsnAlaGlyAspLeuThrVal 1234

3839 .....................ACTCGGCACGGCTTGCCCACGGTGCCGTT 3867
                           :::   :::|||      |||:::
1235 GlyAsnGlyAlaGluIleAsnAlaThrGluGlyAlaAlaThrLeuThrAl 1251

3868 TTCGGGCAAT......ACGGCATCGGCAGGTTCGACATCGGCATCAGCGC 3911
       :::||||||      |||::::::||||||||:::   :::|||::::
1251 aThrGlyAsnThrLeuThrThrGluAlaGlySerSerIleThrSerThrL 1268

3912 GGGCGCGGGTTTTAGTAGCGGCAGCCTTTCAGACGGCATCAGAGGCAAAA 3961
     :::::::|||   :::   |||      |||::::::|||   |||
1268 ysGlyGlnValAspLeuLeuAla......GlnAsnGlySerIleAlaGly 1282

3962 TCCGCCGCCGCGTGCTGCATTACGGCATTCAGGCAAGATACCGCGCAGGT 4011
     |||      |||:::   :::|||               |||      ::
1283 SerIleAsnAlaAlaAsnValThrLeuAsnThrThrGlyThrLeuThrTh 1299

4012 TTCGGCGGATTCGGCATCGAACCGCACATCGGCGCAACGCGCTATTTCGT 4061
     :   |||   |||      :::   ||||||::      :::|||::::
1299 rValAlaGlySerAspIleLysAlaThrSerGlyThrLeuValIleAsnA 1316

4062 CCAAA 4066
     ::|||
1316 laLys 1317
```

seq_name: pir2:A35140

seq_documentation_block:

```
hemolysin A precursor - Proteus mirabilis
C;Species: Proteus mirabilis
C;Date: 03-Aug-1990 #sequence_revision 03-Aug-1990 #text_change 08-Oct-1999
C;Accession: A35140
R;Uphoff, T.S.; Welch, R.A.
J. Bacteriol. 172, 1206-1216, 1990
A;Title: Nucleotide sequencing of the Proteus mirabilis calcium-independent hemolysin ge
A;Reference number: A35140; MUID:90170827
A;Accession: A35140
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1577 <UPH>
A;Cross-references: GB:M30186; NID:g150888; PIDN:AAA25657.1; PID:g150890


alignment_scores:
           Quality:  277.50            Length:   1715
             Ratio:  0.373               Gaps:     88
 Percent Similarity:  43.382   Percent Identity: 19.767

alignment_block:
 US-09-303-518D-653 x A35140   ..

 Align seg 1/1  to: A35140  from: 1  to: 1577

      40 AAAGCCCCTAAAACCGGCCGCATCCGCTTCTCGCCCGCTTACTTAGCCAT 89
         |||         :::||||||:::      |||   |||   :::   ::
       7 LysLeuSerProSerGlyArgLeuAlaAlaSerLeuAlaIleIlePheVa 23

      90 ATGCCTGTCGTTC.........GGCATTCTGCCCCAAGCCCGGGCGGGAC 130
         :   |||:::            ||||||:::|||:::      |||||||
      23 lSerLeuAsnAlaTyrGlyAsnGlyIleValProAsp......AlaGlyH 38

     131 AC.............................................. 132
         ||
      38 isGlnGlyProAspValSerAlaValAsnGlyGlyThrGlnValIleAsn 54

     133 ............ACTTATTTCGGCATCAACTACCAATACTATCGCGACTT 170
                     :::      ||||||:::::::::   |||:::|||||
      55 IleValThrProAsnAsnGluGlyIleSerHisAsnGlnTyrGlnAspPh 71

     171 TGCCGAAAATAAAGGCAAG...TTTGCAGTCGGGGCGAAAGATATTGAGG 217
         |      |||   ||||||      ||||||      :::   :::|||:
      71 e......AsnValGlyLysProGlyAlaValPheAsnAsnAlaLeuGluA 86

     218 TT.......................TACAACAAAAAAGGGGAGTTGGTC 243
         ::                        |||   :::::::::|||
      86 laGlyGlnSerGlnLeuAlaGlyHisLeuAsnAlaAsnSerAsnLeuAsn 102

     244 GGCAAATCGATGACGAAAGCCCCGATGATTGATTTTTCTGTGGTATCGCG 293
         |||::::::          |||   :::|||       :::|||||||||||
     103 GlyGlnAla.........AlaSerLeuIleLeuAsnGluValValSerAr 116

     294 TAAC.................................GGCGTGGCGGCAT 310
         ||||                                |||:::||||||
     116 gAsnProSerPheLeuLeuGlyGlnGlnGluValPheGlyIleAlaAla. 132

     311 TGGCGGGCGATCAATATATTGTGAGCGTGGCACATAAC............ 348
                 :::|||:::::::|||     ::::::
     133 ..........GluTyrValLeuSerAsnProAsnGlyIleThrCysAsp 145

     349 ......GGCGGCTATAACAATGTTGATTTTGGTGCGGAGGGAAGCAATCC 392
               |||      |||:::         :::         :::|||||
     146 GlyCysGlyPheIleAsnThrSerArgSerSerLeuValValGlyAsnPr 162

     393 C.........GATCAGCACCGCTTTTCTTACCAAATTGTGAAAAGAAATA 433
         |         :::   :::      :::|||:::   ::::::::::::|
     162 oLeuPheGluAsnGlyGlnLeuLysGlyTyrSerThrLeuAsnAsnThrA 179

     434 AT...TATAAAGCAGGGACTAACGGCCATCCTTATGGCGGC......GAT 474
         ||      :::   |||   ||||||            |||      |||
     179 snLeuLeuSerLeuGlyLysAsnGlyLeuAsnThrThrGlyLeuLeuAsp 195

     475 TATCATATGCCGCGTTTGCACAAA......TTTGTCACAGATGCAGAACC 518
                 |||||||:::   :::         :::|||   ||||||
     196 LeuIleAlaProArgIleAspSerArgGlyLysIleThrAlaAlaGluIl 212

     519 TGTTGAGATGACC..................................... 531
               :::|||
     212 eSerAlaPheThrGlyGlnAsnThrPheSerGlnHisPheAspIleLeuS 229

     532 ................AGTTATATGGATGGGTGGAAATACGCTGATTTA 564
                         |||   :::|||::::::   ::::::::::::
     229 erSerGlnLysProValSerAlaLeuAspSerTyrPhePheGlySerMet 245

     565 AATAAATACCCTGATCGTGTTCGAATC..............GGAGCAGG 599
         ::::::         |||:::||||||              |||:::||
     246 GlnSer......GlyArgIleArgIleIleAsnThrAlaGluGlySerGl 260

     600 CAGACAATAT.........TGGCGGTCTGATGAAGAC.........GAAC 631
         |   :::            :::   :::|||:::|||         :::
     260 yValLysLeuAlaGlyLysPheThrAlaAspAsnAspLeuSerValLysA 277

     632 CCAATAACCGCGAAAGTTCATATCATATTGCAAGCGCATAT...TCTTGG 678
           :::|||   :::::::::   ::::::          |||   |||:::
     277 laAspAsnIleGlnThrAspSerGlnVal......ArgTyrAspSerTyr 291

     679 CTCGTCGGTGGCAATACCTTTGCACAAAATGGATCAGGTGGTGGCACAGT 728
                 |||:::         ||||||      ||||||   |||||
     292 AspLysAspGlySerGluAsnTyrGlnAsnTyrArgGlyGlyIleThrVa 308

     729 C...AACTTAGGTAGCGAAAAA........................... 747
         |   |||   ||||||::::::
     308 lAsnAsnSerGlySerSerGlnThrLeuThrLysThrGluLeuLysGlyL 325

     748 ...........ATTAAACATAGCCCATATGGTTTTTTACCAACAGGAGGC 786
                     :::      |||   ::::::   :::   ::::::
     325 ysAsnIleThrLeuValAlaSerSerHisAsnGlnIleLysAlaSerAsp 341

     787 TCATTTGGCGAC............AGTGGCTCACCAATGTTTATC..... 819
           :::||||||             :::|||:::   :::   |||
     342 LeuMetGlyAspAspIleThrLeuGlnGlyAlaAspLeuThrIleAspGl 358

     819 ............................................... 819

     358 yLysGlnLeuGlnGlnLysGluThrAspIleAspAsnArgTrpPheTyrS 375

     820 ........TATGATGCCCAAAAGCAAAAGTGGTTAATTAATGGGGTATTG 861
                 ||||||:::   |||:::|||          :::   :::
     375 erTrpLysTyrAspValThrLysGluLys.........GluGlnIleGln 388

     862 CAAACAGGCAACCCCTATATAGGAAAAAGCAATGGCTTCCAGCTAGTTCG 911
         |||   |||:::         :::|||:::|||:::         :::
     389 GlnIleGlySerGlnIleAspAlaLysAsnAsnAlaThrLeuThrAlaTh 405

     912 TAAA...GATTGGTTCTATGAT......GAAATCTTTGCTGGAGATACCC 952
          |||   |||          |||      ::::||   ||||||::::::
     405 rLysGlyAspValThrLeuAspAlaAlaLysIleAsnAlaGlyAsnAsnL 422

     953 ATTCAGTA......TTCTACGAACCACATCAAAATGGG.........AAA 987
           ::::::             :::   |||   ||||||         |||
     422 euAlaIleAsnAlaAsnLysAspIleHisIleAsnGlyLeuValGluLys 438

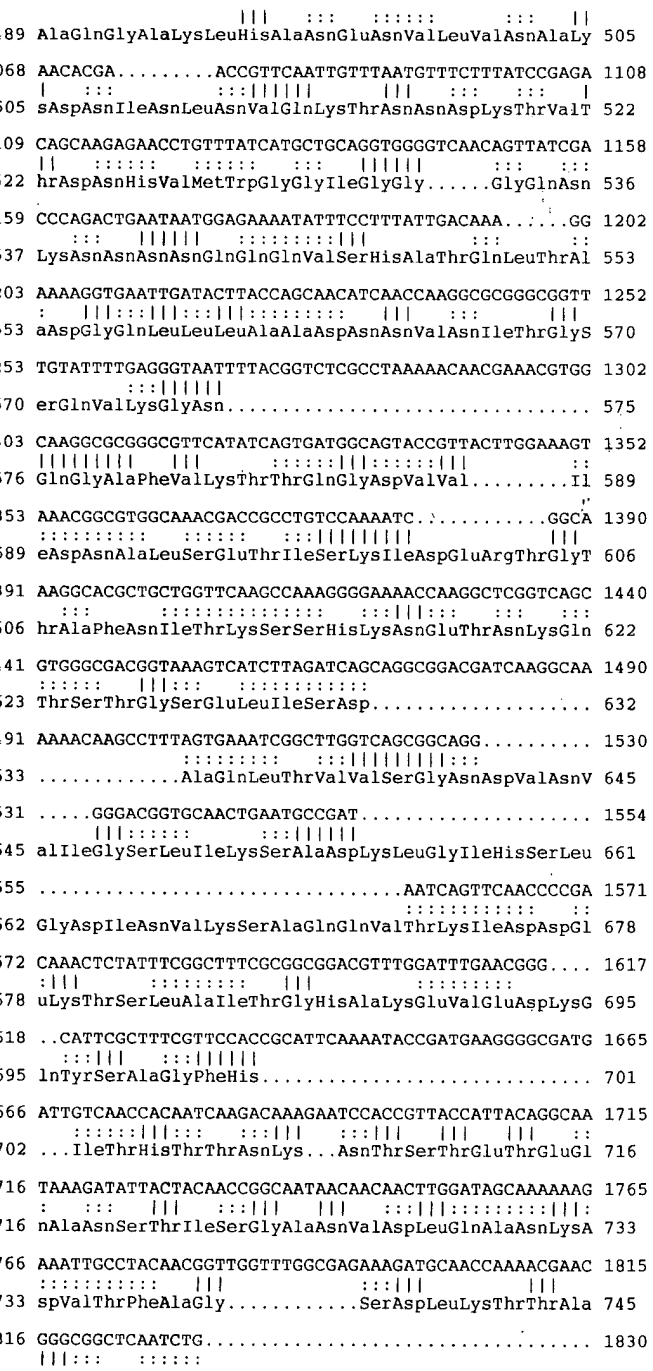
     988 TACTTTTTTAACGACAATAATAAT........................ 1011
                 :::::::|||:::|||
     439 GluSerArgSerGluAsnGlyAsnLysArgAsnHisThrSerArgLeuGl 455

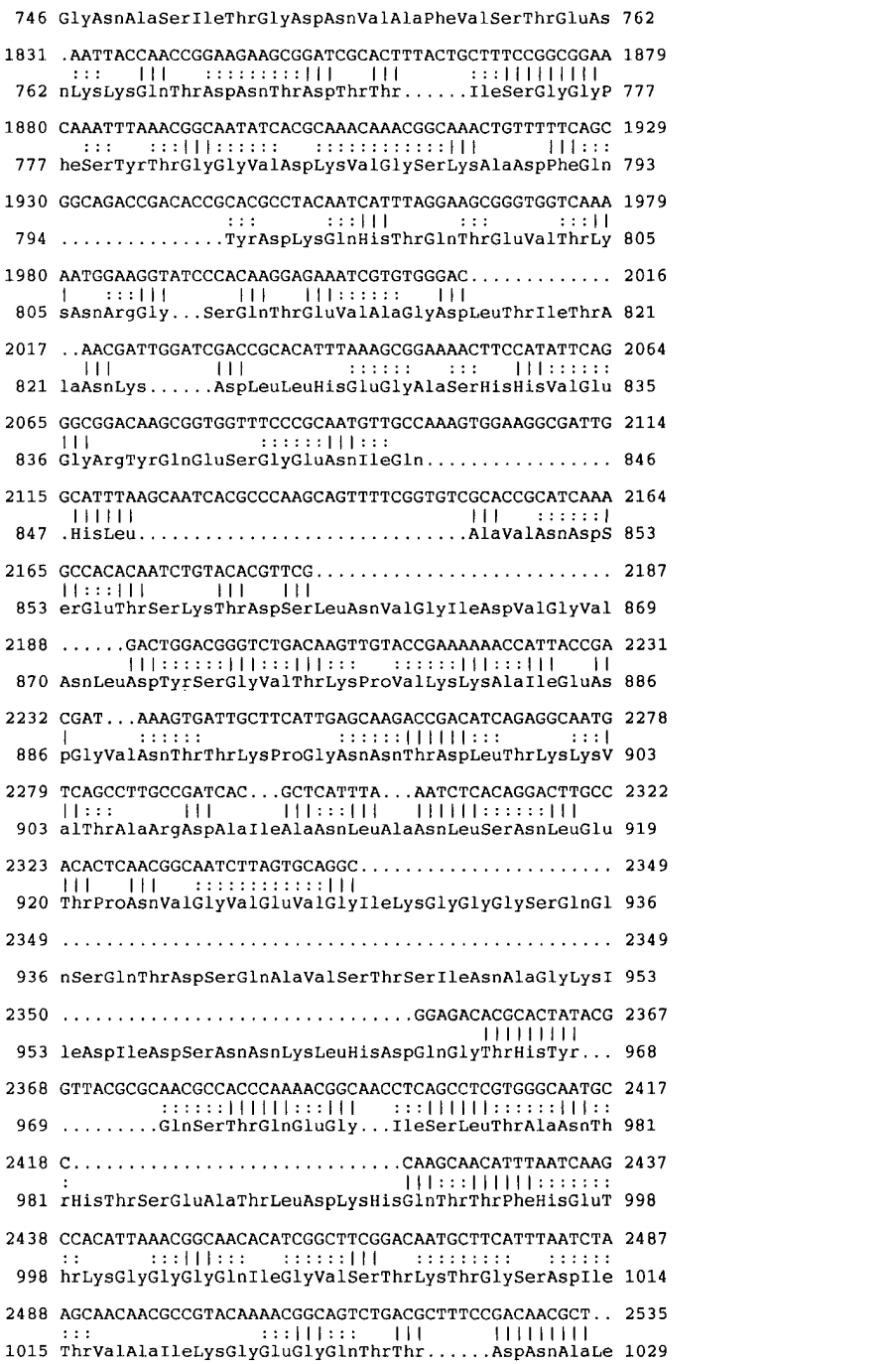
    1011 ............................................... 1011

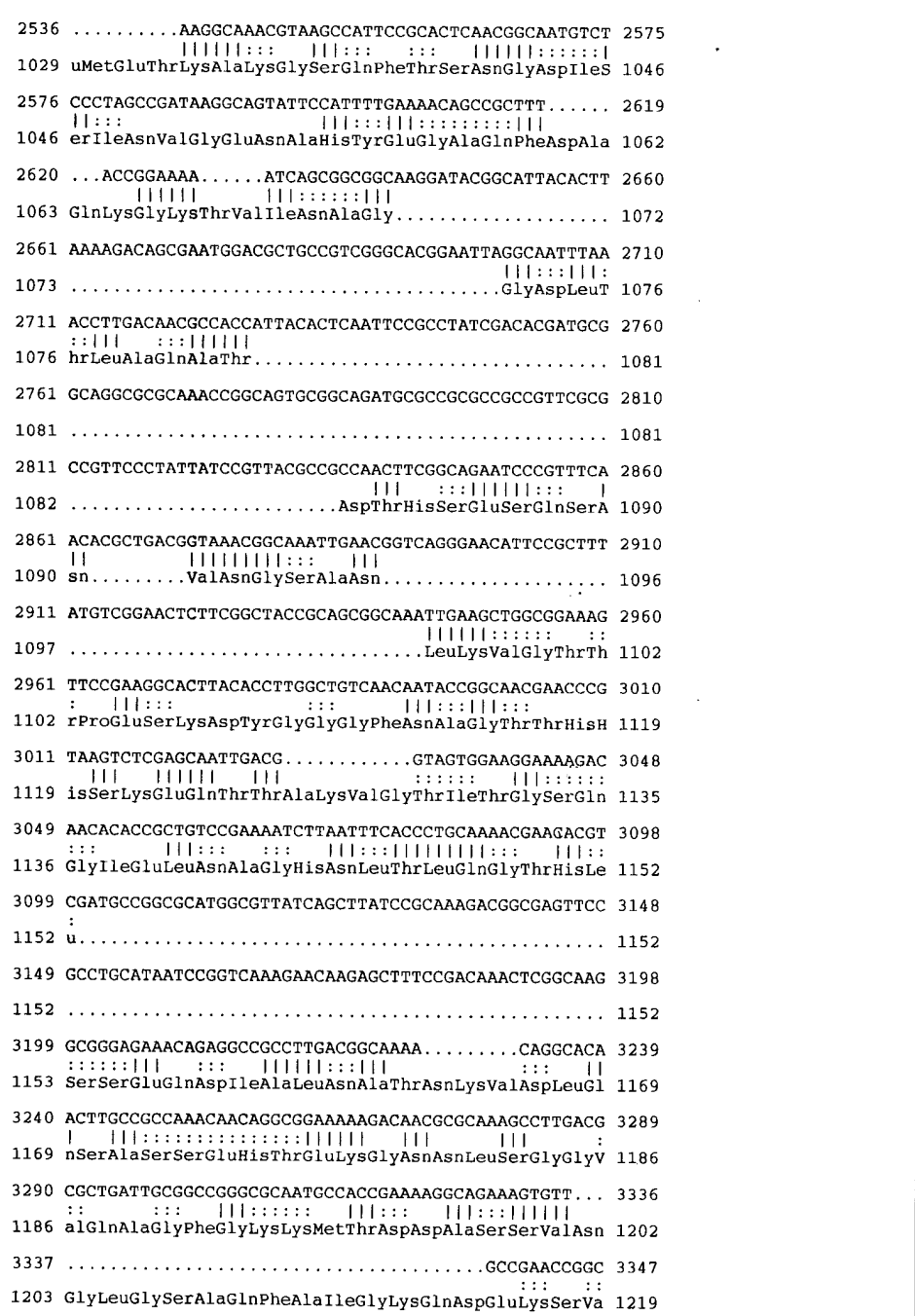
     455 uSerGlySerTrpSerAsnSerHisGlnThrGluThrLeuLysAlaSerG 472

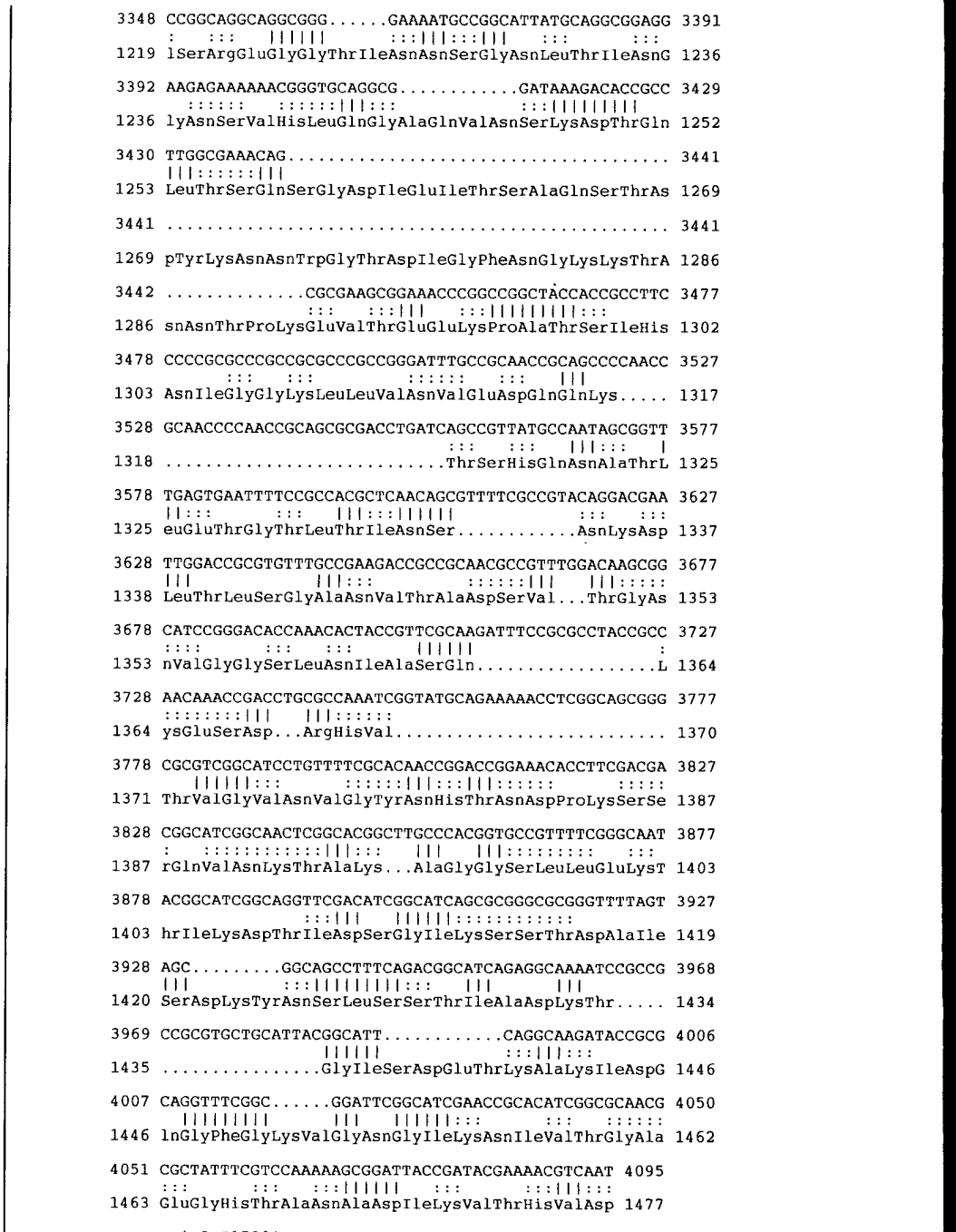
    1012 .....GGCGCAGGAAAA.........ATCGATGCCAAA............ 1035
                 |||||||||         :::|||||||:::
     472 luLeuThrAlaGlyLysAspLeuGlyLeuAspAlaGlnGlySerIleThr 488

    1036 .................CATAAACACTATTCTCTACCTTATAGATTAAA 1067
```


489 AlaGlnGlyAlaLysLeuHisAlaAsnGluAsnValLeuValAsnAlaLy 505
1068 AACACGA.....ACCGTTCAATTGTTTAATGTTTCTTTATCCGAGA 1108
505 sAspAsnIleAsnLeuAsnValGlnLysThrAsnAsnAspLysThrValT 522
1109 CAGCAAGAGAACCTGTTTATCATGCTGCAGGTGGGGTCAACAGTTATCGA 1158
522 hrAspAsnHisValMetTrpGlyGlyIleGlyGly.....GlyGlnAsn 536
1159 CCCAGACTGAATAATGGAGAAAATATTTCCTTTATTGACAAA.....GG 1202
537 LysAsnAsnAsnAsnGlnGlnGlnValSerHisAlaThrGlnLeuThrAl 553
1203 AAAAGGTGAATTGATACTTACCAGCAACATCAACCAAGGCGCGGGCGGTT 1252
553 aAspGlyGlnLeuLeuLeuAlaAlaAspAsnAsnValAsnIleThrGlyS 570
1253 TGTATTTTGAGGGTAATTTTACGGTCTCGCCTAAAAACAACGAAACGTGG 1302
570 erGlnValLysGlyAsn..... 575
1303 CAAGGCGCGGGCGTTCATATCAGTGATGGCAGTACCGTTACTTGGAAAGT 1352
576 GlnGlyAlaPheValLysThrThrGlnGlyAspValVal.....Il 589
1353 AAACGGCGTGGCAAACGACCGCCTGTCCAAAATC.....GGCA 1390
589 eAspAsnAlaLeuSerGluThrIleSerLysIleAspGluArgThrGlyT 606
1391 AAGGCACGCTGCTGGTTCAAGCCAAAGGGGAAAACCAAGGCTCGGTCAGC 1440
606 hrAlaPheAsnIleThrLysSerSerHisLysAsnGluThrAsnLysGln 622
1441 GTGGGCGACGGTAAAGTCATCTTAGATCAGCAGGCGGACGATCAAGGCAA 1490
623 ThrSerThrGlySerGluLeuIleSerAsp..... 632
1491 AAAACAAGCCTTTAGTGAAATCGGCTTGGTCAGCGGCAGG..... 1530
633 .....AlaGlnLeuThrValValSerGlyAsnAspValAsnV 645
1531 .....GGGACGGTGCAACTGAATGCCGAT..... 1554
645 alIleGlySerLeuIleLysSerAlaAspLysLeuGlyIleHisSerLeu 661
1555 .....AATCAGTTCAACCCCGA 1571
662 GlyAspIleAsnValLysSerAlaGlnGlnValThrLysIleAspAspGl 678
1572 CAAACTCTATTTCGGCTTTCGCGGCGGACGTTTGGATTTGAACGGG.... 1617
678 uLysThrSerLeuAlaIleThrGlyHisAlaLysGluValGluAspLysG 695
1618 ..CATTCGCTTTCGTTCCACCGCATTCAAAATACCGATGAAGGGGCGATG 1665
695 lnTyrSerAlaGlyPheHis..... 701
1666 ATTGTCAACCACAATCAAGACAAAGAATCCACCGTTACCATTACAGGCAA 1715
702 ...IleThrHisThrThrAsnLys...AsnThrSerThrGluThrGluGl 716
1716 TAAAGATATTACTACAACCGGCAATAACAACAACTTGGATAGCAAAAAAG 1765
716 nAlaAsnSerThrIleSerGlyAlaAsnValAspLeuGlnAlaAsnLysA 733
1766 AAATTGCCTACAACGGTTGGTTTGGCGAGAAAGATGCAACCAAAACGAAC 1815
733 spValThrPheAlaGly.....SerAspLeuLysThrThrAla 745
1816 GGGCGGCTCAATCTG..... 1830


746 GlyAsnAlaSerIleThrGlyAspAsnValAlaPheValSerThrGluAs 762
1831 .AATTACCAACCGGAAGAAGCGGATCGCACTTTACTGCTTTCCGGCGGAA 1879
762 nLysLysGlnThrAspAsnThrAspThrThr.....IleSerGlyGlyP 777
1880 CAAATTTAAACGGCAATATCACGCAAACAAACGGCAAACTGTTTTTCAGC 1929
777 heSerTyrThrGlyGlyValAspLysValGlySerLysAlaAspPheGln 793
1930 GGCAGACCGACACCGCACGCCTACAATCATTTAGGAAGCGGGTGGTCAAA 1979
794 .....TyrAspLysGlnHisThrGlnThrGluValThrLy 805
1980 AATGGAAGGTATCCCACAAGGAGAAATCGTGTGGGAC..... 2016
805 sAsnArgGly...SerGlnThrGluValAlaGlyAspLeuThrIleThrA 821
2017 ..AACGATTGGATCGACCGCACATTTAAAGCGGAAAACTTCCATATTCAG 2064
821 laAsnLys.....AspLeuLeuHisGluGlyAlaSerHisHisValGlu 835
2065 GGCGGACAAGCGGTGGTTTCCCGCAATGTTGCCAAAGTGGAAGGCGATTG 2114
836 GlyArgTyrGlnGluSerGlyGluAsnIleGln..... 846
2115 GCATTTAAGCAATCACGCCCAAGCAGTTTTCGGTGTCGCACCGCATCAAA 2164
847 .HisLeu.....AlaValAsnAspS 853
2165 GCCACACAATCTGTACACGTTCG..... 2187
853 erGluThrSerLysThrAspSerLeuAsnValGlyIleAspValGlyVal 869
2188 ......GACTGGACGGGTCTGACAAGTTGTACCGAAAAAACCATTACCGA 2231
870 AsnLeuAspTyrSerGlyValThrLysProValLysLysAlaIleGluAs 886
2232 CGAT...AAAGTGATTGCTTCATTGAGCAAGACCGACATCAGAGGCAATG 2278
886 pGlyValAsnThrThrLysProGlyAsnAsnThrAspLeuThrLysLysV 903
2279 TCAGCCTTGCCGATCAC...GCTCATTTA...AATCTCACAGGACTTGCC 2322
903 alThrAlaArgAspAlaIleAlaAsnLeuAlaAsnLeuSerAsnLeuGlu 919
2323 ACACTCAACGGCAATCTTAGTGCAGGC..... 2349
920 ThrProAsnValGlyValGluValGlyIleLysGlyGlyGlySerGlnGl 936
2349 2349
936 nSerGlnThrAspSerGlnAlaValSerThrSerIleAsnAlaGlyLysI 953
2350 .....GGAGACACGCACTATACG 2367
953 leAspIleAspSerAsnAsnLysLeuHisAspGlnGlyThrHisTyr... 968
2368 GTTACGCGCAACGCCACCCAAAACGGCAACCTCAGCCTCGTGGGCAATGC 2417
969 .....GlnSerThrGlnGluGly...IleSerLeuThrAlaAsnTh 981
2418 C.....CAAGCAACATTTAATCAAG 2437
981 rHisThrSerGluAlaThrLeuAspLysHisGlnThrThrPheHisGluT 998
2438 CCACATTAAACGGCAACACATCGGCTTCGGACAATGCTTCATTTAATCTA 2487
998 hrLysGlyGlyGlyGlnIleGlyValSerThrLysThrGlySerAspIle 1014
2488 AGCAACAACGCCGTACAAAACGGCAGTCTGACGCTTTCCGACAACGCT.. 2535
1015 ThrValAlaIleLysGlyGluGlyGlnThrThr.....AspAsnAlaLe 1029





seq_name: pir2:S15204

```
seq_documentation_block:
pertactin - Bordetella parapertussis
N;Alternate names: outer membrane protein P70
C;Species: Bordetella parapertussis
C;Date: 07-Apr-1994 #sequence_revision 07-Apr-1994 #text_change 08-Oct-1999
C;Accession: S15204; S14659
R;Li, L.J.; Dougan, G.; Novotny, P.; Charles, I.G.
Mol. Microbiol. 5, 409-417, 1991
A;Title: P.70 pertactin, an outer-membrane protein from Bordetella parapertussis: clonin
A;Reference number: S15204; MUID:91251771
A;Accession: S15204
A;Molecule type: DNA
A;Residues: 1-922 <LIL>
A;Cross-references: EMBL:X54547; NID:g39761; PIDN:CAA38419.1; PID:g39762
C;Genetics:
A;Gene: prn
C;Keywords: membrane protein


alignment_scores:
            Quality:  275.50              Length:   1162
              Ratio:  0.566                Gaps:     54
 Percent Similarity:  41.910   Percent Identity: 21.687

alignment_block:
 US-09-303-518D-653 x S15204   ..

 Align seg 1/1  to: S15204  from: 1  to: 922

    1366 AACGACCGCCTGTCCAAAATCGGCAAAGGC.............ACGCT 1400
         |||      ||||||:::|||   |||:::               |||
       2 AsnMetSerLeuSerArgIleValLysAlaAlaProLeuArgArgThrTh 18

    1401 GCTGGTTCAAGCCAAAGGG............................. 1419
          |||:::::::|||   |||
      18 rLeuAlaMetAlaLeuGlyAlaLeuGlyAlaAlaProAlaAlaTyrAlaA 35

    1420 .....GAAAACCAAGGCTCGGTCAGCGTGGGCGACGGTAAA...GTCATC 1461
              :::|||||||:::   :::::::::|||:::   :::     . |||
      35 spTrpAsnAsnGlnSerIleIleLysAlaGlyGluArgGlnHisGlyIle 51

    1462 TTAGATCAGCAGGCGGACGATCAAGGCAAAAAACAAGCCTTTAGTGAAAT 1511
             :::|||:::|||      |||   :::   |||   :::
      52 HisIleLysGlnSerAspGlyAlaGlyValArgThrAlaThrGlyThrTh 68

    1512 CGGCTTGGTCAGCGGCAGGGGGACGGTGCAACTGAATGCCGATAATCAGT 1561
              ||||||||||||   :::     :::       :::|||
      68 rIleLysValSerGlyArgGlnAlaGlnGlyValLeuLeuGluAsn.... 83

    1562 TCAACCCCGACAAACTCTATTTCGGCTTTCGCGGCGGACGTTTGGATTTG 1611
                           :::   |||:::::::|||   :::
      84 ........ProAlaAlaGluLeuArgPheGlnAsnGlySerValThrSer 97

    1612 AACGGGCATTCGCTTTCGTTCCACCGCATTCAAAATACCGATGAAGGGGC 1661
         :::|||                          |||      |||||||||::
      98 SerGly.......................GlnLeuPheAspGluGlyVa 106

    1662 GATGATTGTCAACCACAATCAAGACAAAGAATCCACCGTTACCATTACAG 1711
         :          :::          :::||||||||||:::   :
     106 l.................ArgArgPheLeuGlyThrValThrValLysA 117

    1712 GCAATAAA.........GATATTACTACAACCGGCAATAACAACAACTTG 1752
         :::::|||         |||   :::|||   :::|||   ::::::
     117 laGlyLysLeuValAlaAspHisAlaThrLeuAlaAsnValSerAspThr 133

    1753 GATAGCAAAAAAGAAATTGCC...TACAACGGTTGGTTTGGCGAGAAAGA 1799
           :::         ||||||   |||   :::       |||||:::
     134 ArgAspAspAspGlyIleAlaLeuTyrValAla......GlyGluGln.. 147

    1800 TGCAACCAAAACGAACGGGCGGCTCAATCTGAATTACCAACCGGAAGAAG 1849
                                              |||   :::   |
     148 ................................AlaGlnAlaSerIleA 153

    1850 CGGATCGCACTTTACTGCTTTCCGGCGGAACAAATTTA......AACGGC 1893
         |||||   ||||||      :::||||||:::::::::      ::::::
     153 laAspSerThrLeuGlnGlyAlaGlyGlyValArgValGluArgGlyAla 169

    1894 AATATCACGCAAACAAACGGCAAACTGTTTTTCAGCGGCAGACCGACACC 1943
         |||:::|||      ::::::   :::      :::|||
     170 AsnValThrValGlnArgSerThrIleValAspGlyGlyLeu........ 183

    1944 GCACGCCTACAATCATTTAGGAAGCGGGTGGTCA...AAAATGGAAGGTA 1990
                  |||:::|||:::            :::   |||   :
     184 .............HisIleGlyThrLeuGlnProLeuGlnProGluAspL 196

    1991 TCCCACAAGGAGAAATCGTGTGGGACAACGATTGGATCGACCGC...... 2034
         ::|||   ::::::::::|||      :::      :::
     196 euProProSerArgValValLeuGlyAspThrSerValThrAlaValPro 212

    2035 ..........................ACATTTAAAGCGGAAAACTTCCA 2057
                                        :::|||   |||:::::::::
     213 AlaSerGlyAlaProAlaAlaValPheValPheGlyAlaAsnGluLeuTh 229

    2058 TATTCAGGGCGGACAAGCGGTGGTTTCCCGCAATGTTGCCAAAGTGGAAG 2107
          ::::::||||||
     229 rValAspGlyGly..................................... 233

    2108 GCGATTGGCATTTAAGCAATCACGCCCAAGCAGTTTTCGGTGTCGCACCG 2157
                |||::::::::::         |||      |||||||||
     234 ........HisIleThrGlyGlyArgAlaAla......GlyValAla... 244

    2158 CATCAAAGCCACACAATCTGTACACGTTCGGACTGGACGGGTCTGACAAG 2207

     244 ..................................................... 244

    2208 TTGTACCGAAAAAACCATTACCGACGATAAAGTGATTGCTTCATTGAGCA 2257
                                  :::   |||   ::::|||::   |||::::
     245 ..................AlaMetAspGlyAlaIleValHisLeuGlnA 255

    2258 AGACCGACATCAGAGGCAATGTCAGCCTTGCCGATCACGCTCATTTAAAT 2307
         :::::   ||||||
     255 rgAlaThrIleArg................................... 259

    2308 CTCACAGGACTTGCCACACTCAACGGCAATCTTAGTGCAGGCGGAGACAC 2357
                               :::|||:::      |||||||||
     260 ......................ArgGlyAspAlaProAlaGlyGly..... 267

    2358 GCACTATACGGTTACGCGCAACGCCACCCAAAACGGCAACCTC....... 2400
                :::|||      :::||||:::   ::::|||   :::
     268 .......AlaValProGlyGlyAlaValProGlyGlyAlaValProGlyG 282

    2401 .............................................AGC 2403
                                                       |||
     282 lyPheGlyProLeuLeuAspGlyTrpTyrGlyValAspValSerAspSer 298

    2404 CTCGTGGGCAATGCCCAAGCAACATTTAATCAAGCCACATTAAACGGCAA 2453
            |||      ||||||:::      :::          |||::::::
     299 ThrValAspLeuAlaGlnSerIleValGluAlaProGlnLeuGlyAlaAl 315

    2454 CACATCGGCTTCGGACAATGCTTCATTTAATCTAAGC.............. 2490
                 |||:::   :::|||      ::::::|||
     315 aIleArgAlaGlyArgGlyAlaArgValThrValSerGlyGlySerLeuS 332

    2491 ...........AACAACGCCGTACAAAACGGC................... 2511
                     :::|||:::::::::::|||
     332 erAlaProHisGlyAsnValIleGluThrGlyGlyGlyAlaArgArgPhe 348

    2512 ......................AGTCTGACGCTTTCCGACAACGCTAAGGC 2540
                               |||:::||||||:::   :::|||::::|
     349 ProProProAlaSerProLeuSerIleThrLeuGlnAlaGlyAlaArgAl 365

    2541 AAAC.................................GTAAGCCATT 2554
         |:::                                   |||:::   :
     365 aGlnGlyArgAlaLeuLeuTyrArgValLeuProGluProValLysLeuT 382
```



```
 846 alGlnProTyrIleLysAlaSerValLeuGlnGluPheAspGlyAlaGly 862

4231 CGAACGCGCGTCAATACCGCCGTATTGGCGCAGGATTTCGGCAAAACCCG 4280
       :::|||:::|||      :::         ::::::      |||||
 863 ThrValArgThrAsnGlyIleAlaHisArgThrGluLeuArgGlyThrAr 879

4281 CAGTGCGGAATGGGGCGTAAACGCCGAAATCAAAGGTTTCACGCTGTCCC 4330
     |   ||||||   |||                          |||::::
 879 g...AlaGluLeuGly.........................LeuGlyM 886

4331 TCCACGCTGCCGCCGCCAAGGGGCCGCAATTGGAAGCG......CAGCAC 4374
     ::   ||||||   :::::::|||   :::|||   |||      ::::::
 886 etAlaAlaAlaLeuGlyArgGlyHisSerLeuTyrAlaSerTyrGluTyr 902

4375 AGCGCGGGCATCAAATTAGGCTACCGCTGG 4404
     |||   |||   ||||||:::      |||
 903 SerLysGlyProLysLeuAlaMetProTrp 912

seq_name: pir2:A47675

seq_documentation_block:
68K outer membrane protein P.68 pertactin - Bordetella bronchiseptica
C;Species: Bordetella bronchiseptica
C;Date: 19-Dec-1993 #sequence_revision 18-Nov-1994 #text_change 08-Oct-1999
C;Accession: A47675
R;Li, J.; Fairweather, N.F.; Novotny, P.; Dougan, G.; Charles, I.G.
J. Gen. Microbiol. 138, 1697-1705, 1992
A;Title: Cloning, nucleotide sequence and heterologous expression of the protective oute
A;Reference number: A47675; MUID:92407514
A;Contents: CN7531
A;Accession: A47675
A;Status: preliminary
A;Molecule type: nucleic acid
A;Residues: 1-911 <LI1>
A;Cross-references: GB:X54815; GB:S46416; NID:g39396; PIDN:CAA38584.1; PID:g39397
A;Note: sequence extracted from NCBI backbone (NCBIN:113318, NCBIP:113319)


alignment_scores:
           Quality:  274.00            Length:   1149
             Ratio:  0.584               Gaps:     50
 Percent Similarity:  40.818   Percent Identity: 20.540

alignment_block:
 US-09-303-518D-653 x A47675   ..

 Align seg 1/1  to: A47675  from: 1  to: 911

1306 GGCGCGGGCGTTCATATCAGTGATGGCAGTACCGTTACTTGGAAAGTAAA 1355
     ||||||||||||:::   :::   |||:::|||:::      ||||||::
   58 GlyAlaGlyValArgThrAlaThrGlyThrThrIle......LysValSe 72

1356 CGGCGTGGCAAACGACCGCCTGTCCAAAATCGGCAAAGGCACGCTGCTGG 1405
     :|||                    :::   :::::::|||:::||||||
  72 rGly.....................ArgGlnAlaGlnGlyValLeuLeuG 82

1406 TTCAAGCCAAAGGGGAAAACCAAGGCTCGGTCAGCGTGGGCGACGGTAAA 1455
       :::      :::|||          :::         :::|||:::
  82 luAsnProAlaAlaGlu.............LeuArgPheGlnAsnGlySer 94

1456 GTCATCTTAGATCAGCAGGCGGACGATCAAGGCAAAAAACAAGCCTTTAG 1505
     |||      :::   |||      |||:::|||   ::::::   :::
  95 ValThrSerSerGlyGlnLeuPheAspGluGlyValArgArgPheLeu.. 110

1506 TGAAATCGGCTTGGTCAGCGGCAGGGGGACGGTGCAACTGAATGCCGATA 1555
                           |||||||||   ::::::|||
 111 .........................GlyThrValThrValLysAla.... 117

1556 ATCAGTTCAACCCCGACAAACTCTATTTCGGCTTTCGCGGCGGACGTTTG 1605
                                        |||:::|||
 118 ......................................GlyLysLeu 120

1606 GATTTGAACGGGCATTCGCTTTCGTTCCACCGCATTCAAAATACCGATGA 1655

       :::      :::|||:::               :::   |||::
 121 ValAlaAspHisAlaThrLeuAlaAsnValSerAspThrArgAspAspAs 137

1656 AGGGGCGATGATTGTCAACCACAATCAAGACAAAGAATCCACCGTT.... 1701
     :|||      :::         :::::::::   ::::::::::::
 137 pGlyIleAlaLeuTyrValAlaGlyGluGlnAlaGlnAlaSerIleAlaA 154

1702 .....ACCATTACAGGCAATAAAGATATTACTACAACCGGCAATAACAAC 1746
          |||:::   |||         :::   :::      :::   |||
 154 spSerThrLeuGlnGlyAlaGlyGlyValArgValGluArgGlyAlaAsn 170

1747 AACTTGGATAGCAAAAAAGAAATTGCCTACAACGGTTGGTTTGGCGAGAA 1796
             ::::::
 171 ValThrValGlnArg.................................... 175

1797 AGATGCAACCAAAACGAACGGGCGGCTCAATCTGAAT...........T 1834
     :::|||   ::::::|||   |||:::::::::
 176 ....SerThrIleValAspGlyGlyLeuHisIleGlyThrLeuGlnProL 191

1835 ACCAACCGGAAGAAGCGGATCGCACTTTACTGCTTTCCGGCGGAACAAAT 1884
       |||||||||:::         :::   ::::::   |||   |||:::
 191 euGlnProGluAspLeuProProSerArgValValLeuGlyAspThrSer 207

1885 TTAAACGGCAATATCACGCAAACAAACGGCAAACTGTTTTTCAGCGGCAG 1934
     :::::::::      :::   ::::::|||
 208 ValThrAla......ValProAlaSerGly..................... 215

1935 ACCGACACCGCACGCCTACAATCATTTAGGAAGCGGGTGGTCAAAAATGG 1984
     :::|||   |||   :::   :::|||::::::          ::::
 216 ....AlaProAlaAlaValSerValPheGlyAlaAsnGluLeuThrValA 231

1985 AAGGTATCCCACAAGGAGAAATCGTGTGGGACAACGATTGGATCGACCGC 2034
     ::|||
 231 spGly..................................................... 232

2035 ACATTTAAAGCGGAAAACTTCCATATTCAGGGCGGACAAGCGGTGGTTTC 2084
                              ||||||   ||||||:::|||:::
 233 ..................GlyHisIleThrGlyGlyArgAlaAla..... 241

2085 CCGCAATGTTGCCAAAGTGGAAGGCGATTGGCATTTAAGCAATCACGCCC 2134
     :::|||||||   ::::::|||
 242 ....GlyValAlaAlaMetAspGly........................... 248

2135 AAGCAGTTTTCGGTGTCGCACCGCATCAAAGCCACACAATCTGTACACGT 2184

 248 ..................................................... 248

2185 TCGGACTGGACGGGTCTGACAAGTTGTACCGAAAAAACCATTACCGACGA 2234

 248 ..................................................... 248

2235 TAAAGTGATTGCTTCATTGAGCAAGACCGACATCAGAGGCAATGTCAGCC 2284
     :::|||:::   |||:::::::::   ||||||
 249 ....AlaIleValHisLeuGlnArgAlaThrIleArg.............. 259

2285 TTGCCGATCACGCTCATTTAAATCTCACAGGACTTGCCACACTCAACGGC 2334
                                                  :::|||
 260 ..............................................ArgGly 261

2335 AATCTTAGTGCAGGCGGAGACACGCACTATACGGTTACGCGCAACGCCAC 2384
     :::      |||||||||           :::|||      :::|||::
 262 AspAlaProAlaGlyGly............AlaValProGlyGlyAlaVa 274

2385 CCAAAACGGC.....................................A 2395
     :   :::|||                                     :
 274 lProGlyGlyPheGlyProLeuLeuAspGlyTrpTyrGlyValAspValS 291

2396 ACCTCAGCCTCGTGGGCAATGCCCAAGCAACATTTAATCAAGCCACATTA 2445
     ::   |||   |||      ||||||:::      :::          |||
 291 erAspSerThrValAspLeuAlaGlnSerIleValGluAlaProGlnLeu 307

2446 AACGGCAACACATCGGCTTCGGACAATGCTTCATTTAATCTAAGC..... 2490
     ::::::         |||:::   :::|||      ::::::|||
```


308 GlyAlaAlaIleArgAlaGlyArgGlyAlaArgValThrValSerGlyGl 324
2491 .....AACAACGCCGTACAAAACGGC..... 2511
324 ySerLeuSerAlaProHisGlyAsnValIleGluThrGlyGlyGlyAlaA 341
2512 .....AGTCTGACGCTTTCCGACAAC 2532
341 rgArgPheProProProAlaSerProLeuSerIleThrLeuGlnAlaGly 357
2533 GCTAAGGCAAAC.....GT 2546
358 AlaArgAlaGlnGlyArgAlaLeuLeuTyrArgValLeuProGluProVa 374
2547 AAGCCATTCCGCACTCAACGGCAATGTCTCCCTAGCCGATAAGGCAGTAT 2596
374 lLysLeuThrLeuAlaGlyGlyAlaGlnGlyGlnGlyAsp.....IleV 389
2597 TCCATTTTGAAAACAGCCGCTTTACCGGAAAAATCAGCGGCGGCAAGGAT 2646
389 alAlaThrGluLeuProProIleProGlyAlaSerSerGlyProLeuAsp 405
2647 ACGGCATTACACTTAAAAGACAGCGAATGGACGCTGCCGTCGGGCACGGA 2696
406 ValAlaLeu...AlaSerGlnAlaArgTrpThr...GlyAlaThrArgAl 420
2697 ATTAGGCAATTTAAACCTTGACAACGCCACCATTACACTCAATTCCGCCT 2746
420 aValAspSerLeuSerIleAspAsnAlaThrTrpValMetThrAsp.... 435
2747 ATCGACACGATGCGGCAGGCGCGCAAACCGGCAGTGCGGCAGATGCGCCG 2796
436 .....AsnSerAsnValGlyAlaLeu 442
2797 CGCCGCCGTTCGCGCCGTTCCCTATTATCCGTTACGCCGCCAACTTCGGC 2846
443 ArgLeuAlaSerAspGlySer...ValAspPheGlnGlnProAlaGluAl 458
2847 AGAATCCCGTTTCAACACGCTGACGGTAAACGGCAAATTGAACGGTCAGG 2896
458 a...GlyArgPheLysCysLeuMetValAsp...ThrLeuAlaGlySerG 473
2897 GAACATTCCGCTTTATGTCGGAACTCTTCGGCTACCGCAGCGGCAAATTG 2946
473 lyLeuPheArgMetAsnValPheAlaAspLeuGlyLeuSerAspLysLeu 489
2947 AAGCTGGCGGAAAGTTCCGAAGGCACTTACACCTTGGCTGTCAACAATAC 2996
490 ValValMetArgAspAlaSerGlyGlnHisArgLeuLeuValArgAsnSe 506
2997 CGGCAACGAACCCGTAAGTCTCGAGCAATTGACGGTAGTGGAAGGAAAAG 3046
506 rGlySerGluProAlaSerGlyAsnThrMetLeuLeuValGln..... 520
3047 ACAACACACCGCTGTCCGAAAATCTTAATTTCACCCTGCAAAACGAA... 3093
521 .....ThrProArgGlySerAlaAlaThrPheThrLeuAlaAsnLysAsp 535
3094 ...CACGTCGATGCCGGCGCATGGCGTTATCAGCTTATCCGCAAA...GA 3137
536 GlyLysValAspIleGlyThrTyrArgTyrArgLeuAlaAlaAsnGlyAs 552
3138 CGGCGAGTTCCGCCTGCATAATCCGGTCAAAGAACAAGAGCTTTCCGACA 3187
552 nGlyGlnTrpSerLeu.Val..... 558
3188 AACTCGGCAAGGCGGGAGAAACAGAGGCCGCCTTGACGGCAAAACAGGCA 3237
558 558
3238 CAACTTGCCGCCAAACAACAGGCGGAAAAAGACAACGCGCAAAGCCTTGA 3287
558 558
3288 CGCGCTGATTGCGGCCGGGCGCAATGCCACCGAAAAGGCAGAAAGTGTTG 3337
559 .....GlyAlaLysAlaProProAlaProLysProAla 569
3338 CCGAACCGGCCCGGCAGGCAGGCGGGGAAAATGCCGGCATTATGCAGGCG 3387
570 Pro...GlnProGlyProGln..... 575
3388 GAGGAAGAGAAAAAACGGGTGCAGGCGGATAAAGACACCGCCTTGGCGAA 3437
575 575
3438 ACAGCGCGAAGCGGAAACCCGGCCGGCTACCACCGCCTTCCCCCGCGCCC 3487
576 .....ProGlyProGlnProProGln.ProProGlnPr 586
3488 GCCGCGCCCGCCGGGATTTGCCGCAACCGCAGCCCCAACCGCAACCCCAA 3537
586 oProGlnProProGlnArgGlnProGluAla.ProAlaProGlnProPro 602
3538 CCGCAGCGCGACCTGATCAGCCGTTATGCCAATAGCGGTTTGAGTGAATT 3587
603 AlaGlyArgGluLeu...SerAlaAlaAlaAsnAlaAlaValAsnThrGl 618
3588 TTCCGCCACGCTCAACAGCGTT.....TTCGCCGTACAGGACGAATTGG 3631
618 yGlyValGlyLeuAlaSerThrLeuTrpTyrAlaGluSerAsnAlaLeuS 635
3632 ACCGCGTGTTTGCCGAAGACCGCCGCAAC.....GCCGTTTGG 3669
635 erLysArgLeuGlyGluLeuArgLeuAsnProAspAlaGlyGlyAlaTrp 651
3670 ACAAGCGGCATCCGGGACACCAAACACTACCGTTCGCAAGATTTCCGCGC 3719
652 GlyArgGlyPhe.....Al 656
3720 CTACCGCCAACAAACCGACCTGCGCCAAATCGGTATGCAGAAAAACCTCG 3769
656 aGlnArgGlnGlnLeuAsp..... 662
3770 GCAGCGGGCGCGTCGGCATCCTGTTTTCGCACAACCGGACCGGAAACACC 3819
663 .....AsnArgAlaGlyArgArg 668
3820 TTCGACGAC.....GGCATCGGCAACTCGGCACG 3848
669 PheAspGlnLysValAlaGlyPheGluLeuGlyAlaAspHisAlaValAl 685
3849 GCTTGCCCACGGTGCC.....GTTTTCGGGCAATACGGCATCGGCAGGT 3892
685 aValAlaGlyGlyArgTrpHisLeuGlyGlyLeuAlaGlyTyrThrArgG 702
3893 TCGACATCGGCATCAGCGCGGGCGCGGGTTTTAGTAGCGGCAGCCTT... 3939
702 lyAspArgGlyPheThrGlyAspGlyGlyGlyHisThrAspSerValHis 718
3939 3939
719 ValGlyGlyTyrAlaThrTyrIleAlaAsnSerGlyPheTyrLeuAspAl 735
3940 .....TCAG 3943
735 aThrLeuArgAlaSerArgLeuGluAsnAspPheLysValAlaGlySerA 752
3944 ACGGC.....ATCAGAGGCAAAATCCGCCGCCGCGTGCTGCATTACGGC 3987
752 spGlyTyrAlaValLysGlyLysTyrArgThr.....HisGly 764
3988 ATTCAGGCAAGATACCGCGCAGGT.....TTCGGCGGATT 4022
765 ValGlyAlaSerLeuGluAlaGlyArgArgPheAlaHisAlaAspGlyTr 781

```
 4023 CGGCATCGAACCGCACATCGGCGCAACGCGCTATTTCGTCCAAAAAGCGG 4072
      :  :::||||||:::         :::  :::  |||     ::::
  781 pPheLeuGluProGlnAlaGluLeuAlaValPheArgValGlyGlyGlyS 798

 4073 ATTACCGATACGAAAAC...GTCAATATCGCCCACCCCGGGCCTTGCATTC 4119
      ::|||||     |||  :::::::::        |||  :::
  798 erTyrArgAlaAlaAsnGlyLeuArgValArgAspGluGlyGlySerSer 814

 4120 AACCGCTACCGCGCGGGCATTAAGGCAGATTATTCATTCAAACCGGCGCA 4169
             |||  |||:::::::::        :::::  |||
  815 ValLeuGlyArgLeuGlyLeuGluValGlyLysArgIleGluLeuAlaGl 831

 4170 ACACATTTCCATCACGCCTTATTTGAGCCTGTCCTATACCGATGCCGCTT 4219
           ::::::  ||||||::::::  |||     :::     :
  831 yGlyArgGlnValGlnProTyrIleLysAlaSerValLeuGlnGluPheA 848

 4220 CCGGCAAAGTCCGAACGCGCGTCAATACCGCCGTATTGGCGCAGGATTTC 4269
      ::|||       :::|||:::|||     :::       ::::::
  848 spGlyAlaGlyThrValArgThrAsnGlyIleAlaHisArgThrGluLeu 864

 4270 GGCAAAACCCGCAGTGCGGAATGGGGCGTAAACGCCGAAATCAAAGGTTT 4319
           ||||||  ||||||  |||
  865 ArgGlyThrArg...AlaGluLeuGly..................... 872

 4320 CACGCTGTCCCTCCACGCTGCCGCCGCCAAGGGGCCGCAATTGGAAGCG. 4368
        |||::::::  ||||||  ::::::|||  :::|||  |||
  873 ....LeuGlyMetAlaAlaAlaLeuGlyArgGlyHisSerLeuTyrAlaS 888

 4369 .....CAGCACAGCGCGGGCATCAAATTAGGCTACCGCTGG 4404
           ::::::|||  |||  ||||||:::      |||
  888 erTyrGluTyrSerLysGlyProLysLeuAlaMetProTrp 901
```

seq_name: pir2:F90073

seq_documentation_block:
hypothetical protein SA2447 [imported] - Staphylococcus aureus (strain N315)
C;Species: Staphylococcus aureus
C;Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 22-Oct-2001
C;Accession: F90073
R;Kuroda, M.; Ohta, T.; Uchiyama, I.; Baba, T.; Yuzawa, H.; Kobayashi, I.; Cui, L.; Oguc
ma, A.; Mizutani-Ui, Y.; Kobayashi, N.; Sawano, T.; Inoue, R.; Kaito, C.; Sekimizu, K.;
 C.; Shiba, T.; Hattori, M.; Ogasawara, N.; Hayashi, H.; Hiramatsu, K.
Lancet 357, 1225-1240, 2001
A;Title: Whole genome sequencing of meticillin-resistant Stapylococcus aureus.
A;Reference number: A89758; MUID:21311952; PMID:11418146
A;Accession: F90073
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-2271 <KUR>
A;Cross-references: GB:BA000018; PID:g13702612; PIDN:BAB43752.1; GSPDB:GN00149
A;Experimental source: strain N315
C;Genetics:
A;Gene: SA2447

alignment_scores:
            Quality: 274.00   Length: 1551
              Ratio: 0.365   Gaps: 61
 Percent Similarity: 48.420   Percent Identity: 18.311

alignment_block:
 US-09-303-518D-653 x F90073   ..

 Align seg 1/1 to: F90073 from: 1 to: 2271

```
  163 CGCGACTTTGCCGAAAATAAAGGCAAGTTTGCAGTCGGGGCGAAAGATAT 212
      |||  :::  ::::::|||:::          |||  |||:::||
   20 ArgLeuTyrLysSerGlyLysAsnTrpValLysSerGlyIleLysGluIl 36

  213 TGAGGTTTACAACAAAAAAGGGGAGTTGGTCGGCAAATCGATGACGAAAG 262
      ||||:::::::::     |||
   36 eGluMetPheLysIleMetGly.........................L 44

  263 CCCCGATGATTGATTTTTCTGTGGTATCGCGTAACGGC..... 300
       |||:::|||:::  |||:::||||||:::::::::
   44 euProPheIleSerHisSerLeuValSerGlnAspAsnGlnSerIleSer 60

  301 .........................GTGGCGGCATTGGCGGGCGATCA 323
                              ::::::|||:::  |||
   61 LysLysMetThrGlyTyrGlyLeuLysThrThrAlaValIleGlyGlyAl 77

  324 ATATATTGTGAGCGTGGCACATAAC..... 348
      :::  |||::::::  |||:::
   77 aPheThrValAsnMetLeuHisAspGlnGlnAlaPheAlaAlaSerAspA 94

  349 .....................................GGCGGCTAT 357
                                          |||:::
   94 laProLeuThrSerGluLeuAsnThrGlnSerGluThrValGlyAsnGln 110

  358 AACAATGTTGATTTTGGTGCGGAGGGAAGCAATCCCGATCAGCACCGCTT 407
      |||::::::  :::  |||:::  |||:::  |||
  111 AsnSerThrThrIleGluAlaSerThrSerThrAlaAsp..... 123

  408 TTCTTACCAAATTGTGAAAAGAAATAATTATAAAGCAGGGACTAACGGCC 457
       |||  :::::::::|||:::::::::          :::|||:::
  124 .SerThrSerValThrLysAsnSerSerSerValGlnThrSerAsnSerA 140

  458 ATCCTTATGGCGGCGATTATCATATGCCGCGTTTGCACAAATTTGTC... 504
             :::::::::          :::  :::|||  :::
  140 spThrValSerSerGlu...........LysSerGluLysValThrSer 152

  505 ...........ACAGATGCAGAACCTGTTGAGATGACCAGTTATATGGA 542
                 |||:::  :::      ::::::||||||      ::
  153 ThrThrAsnSerThrSerAsnGlnGlnGluLysLeuThrSerThrSerGl 169

  543 TGGGTGGAAATACGCTGATTTAAATAAATACCCTGATCGTGTTCGAATCG 592
      ::::  :::      :::  ::::::      |||          ::::
  169 uSerThrSerSerLysAsnThrThrSerSerSerAspThrLysSerValA 186

  593 GAGCAGGCAGACAATATTGGCGGTCTGATGAAGACGAACCCAATAACCGC 642
      :::::                  |||:::  ::::::|||  |||
  186 laSer..............ThrSerSerThrGluGlnProIleAsnThr 197

  643 GAAAGTTCATATCATATTGCAAGC..... 666
      :::::::::        ||||||
  198 SerThrAsnGlnSerThrAlaSerAsnAsnThrSerGlnSerThrThrPr 214

  666 ..... 666

  214 oSerSerValAsnLeuAsnLysThrSerThrThrSerThrSerThrAlaP 231

  667 ............GCATATTCTTGGCTCGTCGGTGGCAATACCTTTGCA 702
                 ::::::|||  |||  :::  :::||||||||||
  231 roValLysLeuArgThrPheSerArgLeu...AlaMetSerThrPheAla 246

  703 CAAAATGGATCA...........GGTGGTGGCACAGTCAACTTAGGTAG 740
      :::  ::::::             ::::::|||:::::::::::::::
  247 SerAlaAlaThrThrThrAlaValThrAlaAsnThrIleThrValAsnLy 263

  741 CGAAAAAATTAAACATAGCCCATATGGTTTTTTACCAACAGGAGGCTCAT 790
      ::::::::::|||:::          ::::::  |||:::|||:::
  263 sAspAsnLeuLysGln...........TyrMetThrThrSerGlyAsnA 276

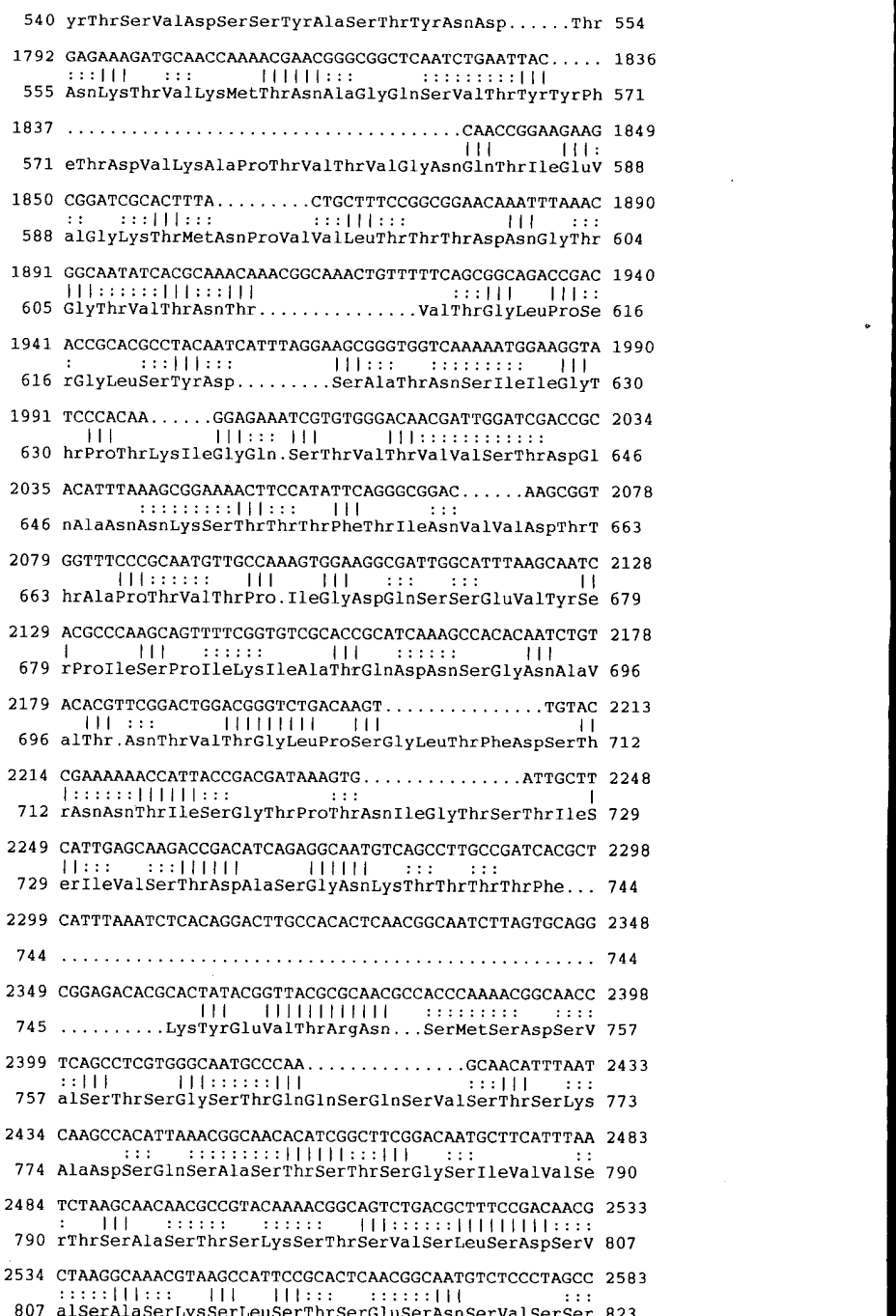
  791 TT.........GGCGACAGTGGCTCACCAATGTTTATCTATGATGCCCAA 831
              :::::::|||      :::      ||||||
  276 laThrTyrAspGlnSerThrGlyIleValThrLeuThrGlnAspAlaTyr 292

  832 AAGCAAAAGTGGTTAATT..... 849
      :::|||||||      |||
  293 SerGlnLysGlyAlaIleThrLeuGlyThrArgIleAspSerAsnLysSe 309

  850 .........AATGGGGTATTGCAAACAGGCAACCCCTATATAGGAAAAA 889
              :::|||  ::::::  ||||||  |||  |||  :
  309 rPheHisPheSerGlyLysValAsnLeuGlyAsnLysTyrGluGlyHisG 326

  890 GCAATGGCTTCCAGCTAGTTCGTAAAGATTGGTTCTATGATGAAATCTTT 939
```


326 lyAsnGly 328
940 GCTGGAGATACCCATTCAGTATTCTACGAACCACATCAAAATGGGAAATA 989
329 ...GlyAspGlyIleGlyPheAlaPheSerProGlyValLeuGlyGluTh 344
990 CTTTTTTAACGACAATAATAATGGCGCAGGA 1020
344 rGlyLeuAsnGlyAlaAlaValGlyIleGlyGlyLeuSerAsnAlaPheG 361
1021 .....AAAATCGATGCCAAACATAAACACTATTCTCTACCTTATAGATTA 1065
361 lyPheLysLeuAspThrTyrHis...AsnThrSerLysProAsn..... 374
1066 AAAACACGAACCGTTCAATTGTTTAATGTTTCTTTATCCGAGACAGCAAG 1115
375 .....SerAlaAlaLysAlaAsnAl 381
1116 AGAACCTGTTTATCATGCTGCAGGTGGGGTCAACAGTTATCGACCCAGAC 1165
381 aAspProSerAsnValAlaGlyGlyGlyAla..... 391
1166 TGAATAATGGAGAAAATATTTCCTTTATTGACAAAGGAAAAGGTGAATTG 1215
392 .....PheGlyAlaPheValThrThrAspSerTyrGlyValAlaThrThr 406
1216 ATACTTACCAGCAACATCAACCAAGGCGCGGGCGGTTTGTATTTTGAGGG 1265
407 TyrThrSerSerSerThrAlaAspAsnAlaAlaLysLeu..... 419
1266 TAATTTTACGGTCTCGCCTAAAAACAACGAAACGTGGCAAGGCGCGGGCG 1315
420 .....AsnValGlnProThrAsnAsn...ThrPheGlnAspPheAspI 433
1316 TTCATATCAGTGATGGCAGTACCGTT..... 1341
433 leAsnTyrAsnGlyAspThrLysValMetThrValLysTyrAlaGlyGln 449
1342 ACTTGGAAAGTAAACGGCGTGGCAAACGACCGCCTGTCCAAAATCGGCAA 1391
450 ThrTrpThrArgAsn......IleSerAspTrpIleAlaLysSerGlyTh 464
1392 AGGCACGCTGCTGGTTCAAGCCAAAGGGGAAAACCAAGGCTCGGTCAGCG 1441
464 rThrAsnPheSerLeuSerMetThrAlaSerThrGlyGlyAlaThrAsnL 481
1442 TGGGCGACGGTAAAGTCATCTTAGATCAGCAGGCGGACGATCAAGGCAAA 1491
481 eu.....GlnGlnVal.....GlnPheGly 487
1492 AAACAAGCCTTTAGTGAAATCGGCTTGGTCAGCGGCAGGGGGACGGTGCA 1541
488 ThrPheGluTyrThrGluSerAlaValThrGlnValArg...TyrValAs 503
1542 ACTGAATGCCGATAATCAGTTCAACCCCGACAAACTCTATTTCGGCTTTC 1591
503 pValThrThrGlyLysAspIleIleProProLysThrTyrSerGly.... 518
1592 GCGGCGGACGTTTGGATTTGAACGGGCATTCGCTTTCGTTCCACCGCATT 1641
518 518
1642 CAAAATACCGATGAAGGGGCGATGATTGTCAACCACAATCAAGACAAAGA 1691
519 ...AsnValAspGlnValValThrIleAspAsn.....GlnGl 530
1692 ATCCACCGTTACCATTACAGGCAATAAAGATATTACTACAACCGGCAATA 1741
530 nSerAlaLeuThrAlaLysGlyTyr.....AsnT 540
1742 ACAACAACTTGGATAGCAAAAAAGAAATTGCCTACAACGGTTGGTTTGGC 1791


540 yrThrSerValAspSerSerTyrAlaSerThrTyrAsnAsp.....Thr 554
1792 GAGAAAGATGCAACCAAAACGAACGGGCGGCTCAATCTGAATTAC..... 1836
555 AsnLysThrValLysMetThrAsnAlaGlyGlnSerValThrTyrTyrPh 571
1837 .....CAACCGGAAGAAG 1849
571 eThrAspValLysAlaProThrValThrValGlyAsnGlnThrIleGluV 588
1850 CGGATCGCACTTTA.....CTGCTTTCCGGCGGAACAAATTTAAAC 1890
588 alGlyLysThrMetAsnProValValLeuThrThrThrAspAsnGlyThr 604
1891 GGCAATATCACGCAAACAAACGGCAAACTGTTTTTCAGCGGCAGACCGAC 1940
605 GlyThrValThrAsnThr.....ValThrGlyLeuProSe 616
1941 ACCGCACGCCTACAATCATTTAGGAAGCGGGTGGTCAAAAATGGAAGGTA 1990
616 rGlyLeuSerTyrAsp.....SerAlaThrAsnSerIleIleGlyT 630
1991 TCCCACAA.....GGAGAAATCGTGTGGGACAACGATTGGATCGACCGC 2034
630 hrProThrLysIleGlyGln.SerThrValThrValValSerThrAspGl 646
2035 ACATTTAAAGCGGAAAACTTCCATATTCAGGGCGGAC.....AAGCGGT 2078
646 nAlaAsnAsnLysSerThrThrThrPheThrIleAsnValValAspThrT 663
2079 GGTTTCCCGCAATGTTGCCAAAGTGGAAGGCGATTGGCATTTAAGCAATC 2128
663 hrAlaProThrValThrPro.IleGlyAspGlnSerSerGluValTyrSe 679
2129 ACGCCCAAGCAGTTTTCGGTGTCGCACCGCATCAAAGCCACACAATCTGT 2178
679 rProIleSerProIleLysIleAlaThrGlnAspAsnSerGlyAsnAlaV 696
2179 ACACGTTCGGACTGGACGGGTCTGACAAGT.....TGTAC 2213
696 alThr.AsnThrValThrGlyLeuProSerGlyLeuThrPheAspSerTh 712
2214 CGAAAAAACCATTACCGACGATAAAGTG.....ATTGCTT 2248
712 rAsnAsnThrIleSerGlyThrProThrAsnIleGlyThrSerThrIleS 729
2249 CATTGAGCAAGACCGACATCAGAGGCAATGTCAGCCTTGCCGATCACGCT 2298
729 erIleValSerThrAspAlaSerGlyAsnLysThrThrThrThrPhe... 744
2299 CATTTAAATCTCACAGGACTTGCCACACTCAACGGCAATCTTAGTGCAGG 2348
744 744
2349 CGGAGACACGCACTATACGGTTACGCGCAACGCCACCCAAAACGGCAACC 2398
745 .....LysTyrGluValThrArgAsn...SerMetSerAspSerV 757
2399 TCAGCCTCGTGGGCAATGCCCAA.....GCAACATTTAAT 2433
757 alSerThrSerGlySerThrGlnGlnSerGlnSerValSerThrSerLys 773
2434 CAAGCCACATTAAACGGCAACACATCGGCTTCGGACAATGCTTCATTTAA 2483
774 AlaAspSerGlnSerAlaSerThrSerThrSerGlySerIleValValSe 790
2484 TCTAAGCAACAACGCCGTACAAAACGGCAGTCTGACGCTTTCCGACAACG 2533
790 rThrSerAlaSerThrSerLysSerThrSerValSerLeuSerAspSerV 807
2534 CTAAGGCAAACGTAAGCCATTCCGCACTCAACGGCAATGTCTCCCTAGCC 2583
807 alSerAlaSerLysSerLeuSerThrSerGluSerAsnSerValSerSer 823

```
2584 GATAAGGCAGTATTCCATTTTGAAAACAGCCGCTTTACCGGAAAAATCAG 2633
 824 SerThrSerThrSerLeuValAsnSerGlnSerValSerSerSerMetSe 840

2634 CGGCGGCAAGGATACGGCATTACACTTAAAAGACAGCGAATGGACGCTGC 2683
 840 rGlySerValSerLysSerThrSerLeuSerAspSerIleSerAsnSerA 857

2684 CGTCGGGCACGGAATTAGGCAATTTAAACCTTGACAACGCCACCATTACA 2733
 857 snSerThrGluLysSerGluSerLeuSerThrSerThrSerAspSerLeu 873

2734 CTCAATTCCGCCTATCGACACGATGCGGCAGGCGCGCAAACC...GGCAG 2780
 874 ArgThrSerThrSerLeuSerAspSerLeuSerMetSerThrSerGlySe 890

2781 TGCGGCAGATGCGCCGCGCCGCCGTTCGCGCCGTTCCCTATTATCCGTTA 2830
 890 rLeuSerLysSerGlnSerLeuSerThrSerIleSerGlySerSerSerT 907

2831 CGCCGCCAACTTCGGCAGAATCCCGTTTCAACACGCTGACGGTAAACGGC 2880
 907 hrSerAlaSerLeuSerAspSerThrSerAsnAlaIleSerThrSerThr 923

2881 AAATTGAACGGTCAGGGAACATTCCGCTTTATGTCGGAACTCTTCGGCTA 2930
 924 SerLeuSerGluSerAlaSer..............................  930

2931 CCGCAGCGGCAAATTGAAGCTGGCGGAAAGTTCCGAAGGCACTTACACCT 2980
 931 .ThrSerAspSerIleSerIleSerAsnSerIleAlaAsnSerGlnSerA 947

2981 TGGCTGTCAACAATACCGGCAACGAACCCGTAAGTCTCGAGCAATTGACG 3030
 947 laSerThrSerLysSerAspSerGlnSerThrSerIleSerLeuSerThr 963

3031 GTAGTGGAAGGAAAAGACAACACACCGCTGTCCGAAAATCTTAATTTCAC 3080
 964 ..........SerAspSerLysSerMetSerThrSer...........  972

3081 CCTGCAAAACGAACACGTCGATGCCGGCGCATGGCGTTATCAGCTTATCC 3130
 972 ..................................................  972

3131 GCAAAGACGGCGAGTTCCGCCTGCATAATCCGGTCAAAGAACAAGAGCTT 3180
 973 .....................................GluSerLeu  975

3181 TCCGACAAACTCGGCAAGGCGGGAGAAACAGAGGCCGCCTTGACGGCAAA 3230
 976 SerAspSerThrSerThrSerGlySerValSerGlySerLeuSer.....  990

3231 ACAGGCACAACTTGCCGCCAAACAACAGGCGGAAAAAGACAACGCGCAAA 3280
 991 .........IleAlaAlaSerGlnSerValSerThrSerThrSerAspS 1004

3281 GCCTT.....................GACGCGCTGATTGCGGCCGGGCGC 3309
1004 erMetSerThrSerGluIleValSerAspSerIleSerThrSerGlySer 1020

3310 AATGCCACCGAAAAGGCAGAAAGTGTTGCCGAACCGGCCCGGCAGGCAGG 3359
1021 LeuSerAlaSerAspSerLysSerMetSerValSerSerSerMetSerTh 1037

3360 CGGGGAAAATGCCGGCATTATGCAGGCGGAGGAAGAGAAAAAACGGGTGC 3409
1037 rSerGlnSerGlySerThrSerGluSerLeuSerAspSerGlnSerThr. 1053

3410 AGGCGGATAAAGACACCGCCTTGGCGAAACAGCGCGAAGCGGAAACCCGG 3459
1054 ..SerAspSerAsp.........SerLysSerLeuSerLeuSerThrSer 1066

3460 CCGGCTACCACCGCCTTCCCCCGCGCCCGCCGCGCCCGCCGGGATTTGCC 3509
1067 GlnSerGlySerThr.SerThrSerThrSerThrSerAlaSerValArgT 1083

3510 GCAACCGCAGCCCCAACCGCAACCCCAACCGCAGCGCGACCTGATCAGCC 3559
1083 hrSerGluSerGlnSerThrSerGlySerMetSerAlaSerGlnSerAsp 1099

3560 GTTATGCCAATAGCGGTTTGAGTGAATTTTCCGCCACGCTCAACAGCGTT 3609
1100 SerMetSerIle.........................SerThrSerPh 1107

3610 TTCGCCGTACAGGACGAATTGGACCGCGTGTTTGCCGAAGACCGCCGCAA 3659
1107 eSerAspSerThrSerAsp..............SerLysSerAlaSerT 1119

3660 CGCCGTTTGGACAA..............GCGGCATCCGGGACACCAAAC 3694
1119 hrAlaSerSerGluSerIleSerGlnSerAlaSerThrSerThrSerGly 1135

3695 ACTACCGTTCGCAAGATTTCCGCGCCTACCGCCAACAAACCGACCTGCGC 3744
1136 SerValSerThrSerThrSerLeuSerThrSerAsnSerGluArgThrSe 1152

3745 CAAATCGGTATGCAGAAAAACCTCGGCAGCGGGCGCGTCGGCATCCTGTT 3794
1152 rThrSerValSerAspSerThrSerLeuSerThrSerGluSerAspSerI 1169

3795 TT......CGCACAACCGGACCGGAAACACCTTCGACGACGGCATCGGCA 3838
1169 leSerGluSerThrSerThrSerAspSerIleSerGluAlaIleSerAla 1185

3839 ACTCGGCACGGCTTGCCCACGGTGCCGTTTTCGGGCAATACGGCATCGGC 3888
1186 SerGluSerThrSerIleSerLeuSerGluSerAsnSerThrSerAspSe 1202

3889 AGGTTCGACATCGGCATCAGCGCGGGCGCGGGTTTTAGTAGCGGCAGCCT 3938
1202 rGluSerGlnSerAlaSerAla.........PheLeuSerGluSerLeuS 1216

3939 TTCAGACGGCATCAGAGGCAAAATCCGCCGCCGCGTGCTGCATTACGGCA 3988
1216 erGluSerThrSerGluSerThrSerGluSerValSerSerSerThrSer 1232

3989 TTCAGGCAAGATACCGCGCAGGTTTCGGCGGATTCGGCATCGAACCGCAC 4038
1233 GluSerThrSerLeuSerAspSerThrSerGluSerGlySerThrSerTh 1249

4039 ATCGGCGCAACGCGCTATTTCGTCCAAAAAGCGGATTACCGATACGAAAA 4088
1249 rSerLeuSerAsnSerThrSerGlySerAlaSerIleSer...ThrSerT 1265

4089 CGTCAATATCGCCACCCCGGGCCTTGCATTCAACCGCTACCGCGCGGGC. 4137
1265 hrSerIleSer..............GluSerThrSerThrPheLysSer 1276

4138 ......ATTAAGGCAGATTATTCATTCAAACCGGCGCAACACATTTCCAT 4181
1277 GluSerValSerThrSerLeuSerMetSerThrSerThrSerLeuSerAs 1293

4182 CACGCCTTATTTGAGCCTGTCCTATACCGATGCCGCTTCCGGCAAAGTCC 4231
1293 nSerThrSerLeuSerThrSerLeuSerAspSerThrSerAspSerLysS 1310

4232 GAACGCGCGTCAATACCGCCGTATTGGCGCAGGAT...TTCGGCAAAACC 4278
1310 erAspSerLeuSerThrSerMetSerThrSerAspSerIleSerThrSer 1326

4279 CGCAGTGCGGAATGGGGCGTAAACGCCGAAATCAAAGGTTTCACG 4323
1327 LysSerAspSerIleSerThrSerThrSerLeuSerGlySerThr 1341
```

seq_name: pir2:C82199

```
seq_documentation_block:
RTX toxin RtxA VC1451 [imported] - Vibrio cholerae (strain N16961 serogroup O1)
C;Species: Vibrio cholerae
C;Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 02-Feb-2001
C;Accession: C82199
R;Heidelberg, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwinn, M.L.; Dodson, R.J.;
chardson, D.; Ermolaeva, M.D.; Vamathevan, J.; Bass, S.; Qin, H.; Dragoi, I.; Sellers, P
l, R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.
Nature 406, 477-483, 2000
A;Title: DNA Sequence of both chromosomes of the cholera pathogen Vibrio cholerae.
A;Reference number: A82035; MUID:20406833
A;Accession: C82199
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-4558 <HEI>
A;Cross-references: GB:AE004223; GB:AE003852; NID:g9655942; PIDN:AAF94608.1; GSPDB:GN001
A;Experimental source: serogroup O1; strain N16961; biotype El Tor
C;Genetics:
A;Gene: VC1451
A;Map position: 1


alignment_scores:
            Quality:  273.50              Length:   1747
              Ratio:  0.351                Gaps:     82
 Percent Similarity:  44.591   Percent Identity: 20.034

alignment_block:
 US-09-303-518D-653 x C82199   ..

 Align seg 1/1  to: C82199  from: 1  to: 4558

      37 CGCAAAGCCCCTAAAACCGGCCGCATCCGC...TTCTCGCCCGCTTACTT 83
         |||:::   |||:::||||||:::::::::   :::|||   |||:::
     365 ArgAsnAspProGlnThrGlyGluLeuLysTyrTyrSerThrAlaTrp.. 380

      84 AGCCATATGCCTGTCGTTCGGCATTCTGCCCCAAGCCCGGGCGGGACACA 133

     380 ................................................. 380

     134 CTTATTTCGGCATCAACTAC.........CAATACTATCGCGACTTTGCC 174
           |||       :::|||:::                :::|||::::::
     381 ..TyrLysGluValAsnHisLeuSerAsnLeuAlaAsnGlnAspIleSer 396

     175 GAAAATAAAGGCAAGTTTGCAGTC.........GGGGCG.........AA 206
         :::|||   |||       ||||||         ||||||         ::
     397 AspAsnGlyGlyPheThrAlaValAsnIleAsnGlyAlaTyrThrLeuSe 413

     207 AGATATTGAGGTTTACAACAAAAAAGGG.........GAGTTGGTCGGCA 247
         :|||:::::|||   ::::::::::::          :::   |||   |
     413 rAspLeuLysValGluHisGlnGlnSerValThrValHisAlaValGluL 430

     248 AATCGATGACGAAAGCCCCGATGATTGATTTTTCTGTGGTATCGCGTAAC 297
         |||||:::|||:::         :::   ::::::          |||
     430 ysSerLeuThrGluTyrGluTrpValThrTyrAla............Asn 442

     298 GGCGTGGCGGCATTGGCGGGCGATCAATATATTGTGAGCGTGGCACATAA 347
         |||   |||::::::   :::   :::      :::|||   |||
     443 Gly...AlaValIleAspAlaLysGluValSerLeuSerAspAlaLysMe 458

     348 CGGCGGCTAT.................AACAATGTTGATTTTGGTGCGG 379
           ||||||:::                ::::::||||||      |||
     458 tGlyGlyHisAlaIleTyrAlaAspGlyThrLysValAspValLysAlaV 475

     380 AGGGAAGCAATCCCGATCAGCACCGCTTTTCTTACCAAATTGTG...... 423
           ||||||   :::   :::   :::   |||      |||
     475 alLysSerAsnArgGlnProAsnThrTyrIleTyrAlaLysValLeuGly 491

     423 ................................................. 423

     492 ProTyrThrLysIleValValValGluLeuAlaAsnAspProGluThrGl 508

     424 ..............AAAAGAAATAATTATAAAGCAGGG.......... 447
                         |||:::   ||||||   |||
     508 yAlaLeuLysTyrGlnAlaArgSerTrpTyrLysGluGlyAspHisThrA 525

     448 ...............................ACTAACGGCCATCCTTAT 465
                                         |||:::   :::|||
     525 laAsnIleAlaAsnGlnAspIleSerSerAlaThrGlyTyrAsnProMet 541

     466 GGC...GGCGATTATCATATGCCGCGTTTGCACAAATTTGTCACAGATGC 512
         |||   |||   |||   :::      ||||||      |||:::   ::
     542 GlyLysGlyGlyTyrSerLeuSerAspLeuHisTyrSerValAsnAlaVa 558

     513 AGAACCTGTTGAGATGACCAGTTATATGGATGGGTGGAAATACGCTGATT 562
         ::::   ::::::   |||                        ||||||:
     558 lArgSerThrSerGluThr.....................ValAlaAspI 568

     563 TAAATAAATACCCTGATCGTGTTCGAATCGGAGCAGGCAGACAATATTGG 612
         ::::::::|||   |||::::::   :::
     568 leGluGluTyrThrAspGlnThrLeuPhe..................... 577

     613 CGGTCTGATGAAGACGAACCCAATAAC......CGCGAAAGTTCATATCA 656
                     :::|||   |||         |||||||||
     578 ..............LysProAlaAsnAspSerGlyGluSerSerGlyAs 589

     657 TATTGCAAGCGCATATTCTTGGCTCGTCGGTGGCAATACCTTTGCACAAA 706
           :::         ::::::         |||||||||::::::   :::|
     589 pVal......ArgPheAsnGlyAlaGlyGlyGlyAsnValIleLysSerA 604

     707 AT.....................GGATCAGGTGGTGGCACAGTCAACTTA 735
         ||                     :::|||||||||   :::|||:::
     604 snValThrArgGlyAsnValHisPheAsnGlyGlyGlyIleAlaAsnVal 620

     736 GGTAGCGAAAAAATTAAACAT...AGCCCATATGGTTTTTTACCAACAGG 782
                     |||   |||   |||   :::|||            ::
     621 ............IleLeuHisSerSerGlnPheGly............As 629

     783 AGGCTCATTTGGCGACAGTGGCTCACCAATGTTTATCTATGATGCCCAAA 832
         :   :::|||:::   :::|||:::         |||      :::   :
     629 nThrGluPheAsnGlyGlyGlyAlaAlaAsnValIleValLysSerGlyG 646

     833 AGCAAAAGTGGTTAATTAATGGGGTATTGCAAACAGGCAACCCCTATATA 882
         :::::      |||                        :::|||      :::
     646 luGluGlyAspLeuThrPheArgGlyAlaGlyLeuAlaAsnValLeuVal 662

     883 GGAAAAAGCAATGGCTTCCAGCTAGTTCGTAAAGATTGGTTCTATGATGA 932
            :::|||:::         ::::::                     ::
     663 HisGlnSerGluGlnGlyLysMet.....................As 671

     933 AATCTTTGCTGGAGATACCCATTCAGTATTCTACGAACCACATCAAAATG 982
         :::::::||||||   :::   :::|||:::   :::         :::|
     671 pValTyrAlaGlyGlyAlaValAsnValLeuValArg...LeuGlyAspG 687

     983 GGAAATACTTTTTTAAC.....................GACAATAATAAT 1011
         ||:::|||:::   :::                     :::   ::::::
     687 lyGlnTyrLeuAlaHisLeuLeuAlaTyrGlyAsnIleSerValGlnLys 703

    1012 GGCGCAGGAAAAATCGATGCCAAACATAAACACTATTCTCTACCTTATAG 1061
         |||:::|||
     704 GlySerGly..................................... 706

    1062 ATTAAAAACACGAACCGTTCAATTG......TTTAATGTTTCTTTATCCG 1105
               :::|||:::|||:::|||      :::|||:::      :::
     707 ....AspSerArgValValMetLeuGlyGlyTyrAsnThrHisThrGlnI 722

    1106 AGACAGCAAGAGAACCTGTTTATCATGCTGCAGGTGGGGTCAAC...... 1149
           :::   :::   ::::::   ||||||||||||   |||
     722 leGlySerGlyAsnGlyLeuTrpLeuAlaAlaGlyGlyPheAsnValMet 738

    1150 .................AGTTATCGACCCAGACTGAATAATGGAGAAAA 1181
                           :::            |||   :::|||   ||
     739 ThrGlnValGlyLysGlyAspValAlaAlaValLeuAlaGlyGlyAlaAs 755

    1182 TATTTCCTTTATTGACAAAGGAAAAGGTGAATTGATACTTACCAGCAACA 1231
```

```
 755 nVal...LeuThrLysMetGlyGluGlyGlu......LeuThrSerGlyM 769

1232 TCAACCAAGGCGCGGGCGGTTTGTATTTTGAGGGTAAT..... 1269
 769 etLeuGlyGlyAlaAsnValIleThrHisIleSerAsnAspAspGlnLeu 785

1270 ..............TTTACGGTCTCGCCTAAAAACAACGAAACGTGGCA 1304
 786 SerAsnThrThrAlaValAlaLeuGlyGlyAlaAsnIleLeuThrLysLy 802

1305 AGGCGCGGGC.....................................G 1315
 802 sGlyLysGlyAsnThrLeuAlaValMetGlyGlyGlyAlaAsnValLeuT 819

1316 TTCATATCAGTGATGGCAGTACCGTTACTTGGAAAGTAAACGGCGTGGCA 1365
 819 hrHisValGlyAspGlyThrThrThrGlyValMetValGlyGlyAla... 834

1366 AACGACCGCCTGTCCAAAATCGGCAAAGGCACGCTGCTGGTTCAAGCCAA 1415
 835 ...AsnIleLeuThrLysValGlyAsnGlyAspThrThrGlyIleLeuLe 850

1416 AGGGGAAAACCAAGGCTCGGTCAGCGTGGGCGACGGTAAAGTCATCTTAG 1465
 850 uGlyValGlyAsnValLeuThrHisValGlyAspGlyGlnThrLeu.... 865

1466 ATCAGCAGGCGGACGATCAAGGCAAAAAACAAGCCTTTAGTGAAATCGGC 1515
 866 ...................GlyValMetGlyAlaAlaGlyAsnIlePhe 875

1516 TTGGTCAGCGGCAGGGGGACG.........GTGCAACTGAATGCCGATAA 1556
 876 ThrLysValGlyAspGlyThrSerIleAlaValMetIleGlyAlaGlyAs 892

1557 TCAGTTCAACCCCGACAAACTCTATTTCGGCTTTCGCGGCGGACGTTTGG 1606
 892 nIlePheThr............HisValGlyGluGlyAsnAlaTrpAlaL 905

1607 ATTTGAACGGGCATTCGCTTTCGTTCCACCGCATTCAAAATACCGATGAA 1656
 905 euMetGlyGlyLeuGlyAsnValPheThrLysValGlyAsnGlyAspAla 921

1657 GGGGCGATGATTGTC.................AACCACAATCAAGACAA 1688
 922 LeuAlaLeuMetValAlaGluAlaAsnValPheThrHisIleGlyAspGl 938

1689 AGAATCCACCGTTACCATTACAGGCAATAAAGATATTACTACA...ACCG 1735
 938 yMetSerValAlaLeuMetLeuAlaLysGlyAsnValAlaThrLysValG 955

1736 GCAATAAC.....................AACAACTTGGATAGCAAAAAA 1764
 955 lyAsnGlyThrThrLeuAlaAlaMetValGlyAsnValAsnIlePheThr 971

1765 GAAATTGCCTACAACGGTTGGTTT............GGCGAGAAAGAT.. 1800
 972 HisIleGlyHisGlySerThrPheAlaAlaMetIleGlyGlnAlaAsnIl 988

1801 .GCAACCAAAACGAACGGGCGGCTCAATCTGAATTACCAACCGGAAGAAG 1849
 988 eMetThrLysValGlyAsnAspLeuThrAlaAlaLeuMetValGlyLysA 1005

1850 CGGATCGCACTTTACTGCTTTCCGGCGGAACAAATTTAAAC......... 1890
1005 laAsnIleMetThrHisValGlyAspGlyThrSerLeuGlyLeuPheAla 1021

1891 GGCAATATCACGCAAACA............AACGGCAAACTGTTTTTCAG 1928
1022 GlyGluValAsnValMetThrLysValGlyAsnGlyThrThrLeuAlaAl 1038

1929 CGGCAGACCGACACCGCACGCCTACAATCATTTAGGAAGCGGGTGGTCAA 1978
```

```
1038 aMetPheGlyLysAlaAsnIleMetThrHisValGlyAspGlyLeuThrG 1055

1979 AAATGGAAGGTATCCCACAAGGAGAAATCGTGTGGGACAACGATTGGATC 2028
1055 lyValLeuAlaLeuGlyGluAlaAsnIleVal..................  1065

2029 GACCGCACATTTAAAGCGGAAAACTTCCATATTCAGGGCGGACAAGCGGT 2078
1066 ......ThrLysLeuGlyAspAspPhe......MetGlyValValAlaAl 1078

2079 GGTTTCCCGCAATGTTGCCAAAGTGGAAGGCGAT................ 2112
1078 aAlaLysAlaAsnValValThrHisValGlyAspAlaThrThrAlaAlaV 1095

2112 ..................................................  2112

1095 alLeuAlaGlyLysGlyAsnIleLeuThrLysValGlyGluGlyThrThr 1111

2113 .......................TGGCATTTAAGCAATCACGCC..... 2133
1112 ValGlyLeuLeuIleSerAspValGlyAsnValMetThrHisValGlyAs 1128

2134 .CAAGCAGTTTTCGGTGTCGCACCGCATCAAAGCCACACAATCTGTACAC 2182
1128 pGlyThrThrIleGlyIleAlaLysGlyLysAlaAsnLeuIleThrLysV 1145

2183 GTTCGGACTGGACGGGTCTGACAAGTTGTACCGAAAAAACCATTACCGAC 2232
1145 alGlyAspGlyLeuGlyVal............AsnValThrTrpGlyGln 1157

2233 GATAAAGTGATTGCTTCATTGAGCAAGACCGAC.................. 2265
1158 AlaAsnValPheThrGlnValGlyAspGlyAspArgTyrAsnPheAlaLy 1174

2265 ..................................................  2265

1174 sGlyGluAlaAsnLeuIleThrLysValGlyAspGlyGlnGluValSerV 1191

2266 ..ATCAGAGGCAATGTCAGCCTTGCCGATCACGCT.............. 2298
1191 alValGlnGlyGluAlaAsnIleIleThrHisValGlyAsnGlyAspAsp 1207

2299 CATTTAAATCTCACAGGACTTGCCACACTC..................... 2328
1208 TyrThrGlyAlaTrpGlyLysAlaAsnValIleThrLysValGlyHisGl 1224

2328 ..................................................  2328

1224 yGlnAsnValValLeuAlaLysGlyGluAlaAsnIleValThrGlnValG 1241

2329 ..............................AACGGCAATCTTAGTGCA 2346
1241 lyAspGlyAspSerPheAsnAlaLeuTrpSerLysGlyAsnIleValThr 1257

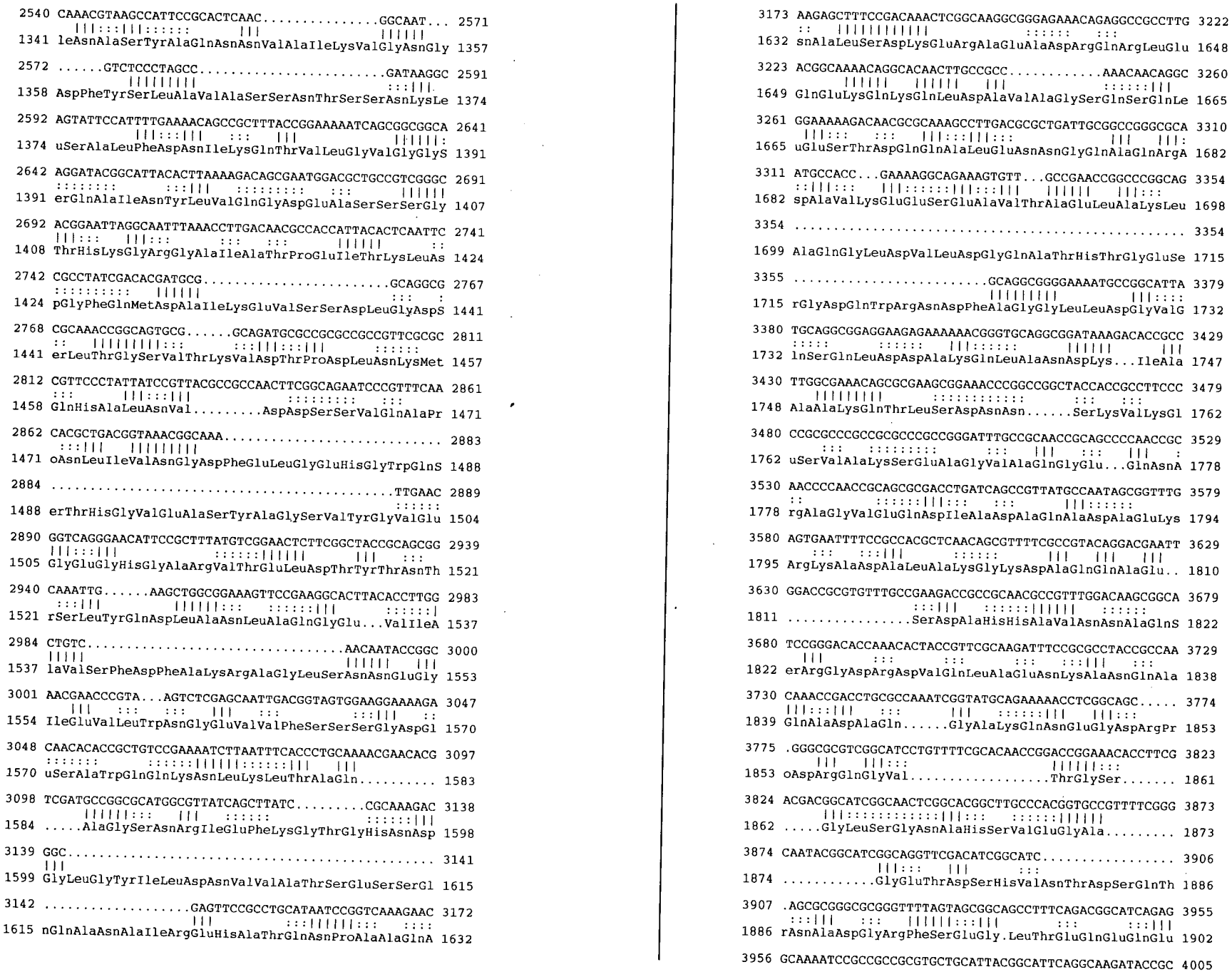
2347 GGCGGAGACACGCACTATACGGTTACGCGCAACGCCACCCAAAACGGCAA 2396
1258 LysValGlyAspGlyMetGlnValThrAlaAlaLysGlyGlnAlaAsnIl 1274

2397 CCTCAGCCTCGTGGGCAATGCC...CAAGCAACATTTAATCAAGCCACAT 2443
1274 eThrThrThrValGlyAsnGlyLeuAsnValThrAlaAlaTyrGlyAspA 1291

2444 TAAACGGCAACACATCGGCTTCGGACAATGCTTCATTTAATCTAAGC... 2490
1291 laAsnIleAsnThrLysValGlyAspGlyValSerValAsnValAlaTrp 1307

2491 ..................................AACAACGCCGT 2501
1308 GlyLysTyrAsnIleAsnThrLysValGlyAspGlyLeuAsnValAlaVa 1324

2502 ACAA............AACGGCAGTCTGACGCTTTCCGACAACGCTAAGG 2539
1324 lMetLysGlyLysAlaAsnAlaAsnIleHisValGlyAspGlyLeuAsnI 1341
```



```
     |||   :::::::|||:::   :::::::::   ::::::   ::::::::
  1903 AlaLeuGluGlyAlaThrAsnAlaValAsnArgLeuGlnIleAsnAlaGl 1919

  4006 GCAGGTTTCGGCGGATTCGGCATCGAACCGCACATCGGCGCAACGCGCTA 4055
       :      :::   :::|||:::|||:::   ||||||
  1919 yIleArgAlaLysAsnSerValSerSerMetThrSerMetPheSerGluT 1936

  4056 TTTCGTCCAAAAAGCGGATTACCGATACGAAAACGTCAATATCGCCACCC 4105
         :::||||||:::   ::::::   ||||||:::|||   :::|||
  1936 hrAsnSerLysSerIleValValProThrLysValSerProGluProGlu 1952

  4106 CGGGCCTTGCATTCAACCGCTACCGCGCGGGCATTAAGG 4144
       |||         :::   :::|||   |||   :::|||
  1953 Arg.........GlnGluValThrArgArgAspValArg 1962
```

seq_name: pir2:B89921

seq_documentation_block:
hypothetical protein ebhA [imported] - Staphylococcus aureus (strain N315)
C;Species: Staphylococcus aureus
C;Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 22-Oct-2001
C;Accession: B89921
R;Kuroda, M.; Ohta, T.; Uchiyama, I.; Baba, T.; Yuzawa, H.; Kobayashi, I.; Cui, L.; Oguc
ma, A.; Mizutani-Ui, Y.; Kobayashi, N.; Sawano, T.; Inoue, R.; Kaito, C.; Sekimizu, K.;
 C.; Shiba, T.; Hattori, M.; Ogasawara, N.; Hayashi, H.; Hiramatsu, K.
Lancet 357, 1225-1240, 2001
A;Title: Whole genome sequencing of meticillin-resistant Stapylococcus aureus.
A;Reference number: A89758; MUID:21311952; PMID:11418146
A;Accession: B89921
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-6713 <KUR>
A;Cross-references: GB:BA000018; PID:g13701232; PIDN:BAB42527.1; GSPDB:GN00149
A;Experimental source: strain N315
C;Genetics:
A;Gene: ebhA

alignment_scores:
            Quality: 273.00   Length: 1667
              Ratio: 0.351   Gaps: 84
 Percent Similarity: 46.611   Percent Identity: 20.756

alignment_block:
 US-09-303-518D-653 x B89921   ..

 Align seg 1/1 to: B89921 from: 1 to: 6713

```
  193 GCAGTCGGGGCGAAAGATATTGAGGTTTACAACAAAAAAGGGGAGTTGGT 242
      |||:::          :::|||   |||   |||:::   |||
  306 AlaIleThrGlnAlaThrThrGlnValAsnAsnAlaLysAsnGlyLeuAs 322

  243 CGGC.....................AAATCGATGACGAAAGCCCCGATGA 271
       |||                    ::::::   :::|||      :::
  322 nGlyAlaGluAsnLeuArgAsnAlaGlnAsnThrAlaLysGlnAsnLeuA 339

  272 TTGATTTTTCTGTGGTATCGCGTAACGGCGTGGCGGCATTG......... 312
          :::|||   ::::::::::|||      :::|||:::
  339 snThrLeuSerHisLeuThrAsnAsnGlnLysSerAlaIleSerSerGln 355

  313 .........GCGGGCGATCAATATATTGTGAGCGTGGCACATAACGGCGG 353
              ||||||           |||:::::::||   |||:::::
  356 IleAspArgAlaGlyHisValSerGluValThrAlaAlaLysAsnAlaAl 372

  354 CTATAACAATGTTGATTTTGGTGCGGAGGGAAGCAATCCCGATCAGCACC 403
      :          ::::::::::::|||:::   :::|||   :::|||
  372 a.........ThrGluLeuAsnAlaGlnMetGlyAsnLeuGluGln.... 384

  404 GCTTTTCTTACCAAATTGTGAAAAGAAATAATTATAAAGCAGGGACTAAC 453
                |||      :::|||:::   |||   |||:::|||
  385 ...........AlaIleHisAspGlnAsnThrValLysGlnGlyValAsn 397

  454 GGCCATCCTTATGGCGGCGATTATCATATGCCGCGTTTGCACAAATTTGT 503

  398 ..............................................Ph 398

  504 CACAGATGCAGAACCTGTTGAGATGACCAGTTATATGGATGGGTGGAAAT 553
       |||||||||:::   ::::::      :::|||   ::::::   :::
  398 eThrAspAlaAspLysAlaLysArgAspAlaTyrThrAsnAlaValSerA 415

  554 ACGCTGAT......TTAAATAAATACCCTGATCGTGTTCGAATCGGAGCA 597
        |||:::      |||||||||
  415 rgAlaGluThrIleLeuAsnLysThrGlnGly..................... 425

  598 GGCAGACAATATTGGCGGTCTGATGAAGACGAACCCAATAACCGCGAAAG 647
                          ::::::   ::::::   :::   :::   ::
  426 ..................AlaAsnThrSerLysGlnAspValGluAlaAl 436

  648 TTCATATCATATTGCAAGCGCATATTCTTGGCTCGTCGGTGGCAAT.... 693
      :      :::::::::||||||   :::   |||   |||   :::
  436 aIleGlnAsnValThrSerAlaLysAsnAlaLeuAsnGlyAspGlnAsnV 453

  694 ..ACCTTTGCACAAAATGGATCAGGTGGTGGCACAGTCAACTTAGGTAGC 741
        |||   |||:::|||::::::   ::::::      ||||||   |||
  453 alThrAsnAlaLysAsnAlaAlaLysAsnAlaLeuAsnAsnLeuThrSer 469

  742 GAAAAAATTAAACATAGCCCATATGGTTTTTTACCAACAGGAGGCTCATT 791

  469 ..............................................  469

  792 TGGCGACAGTGGCTCACCAATGTTTATCTATGATGCCCAAAAGCAAAAGT 841
                                  |||   :::||||||||||:::
  470 .........................IleAsnAsnAlaGlnLysArgAspL 478

  842 GGTTA.................................ATTAATGGGGTA 858
                                          :::::::::|||
  478 euThrThrLysIleAspGlnAlaThrThrValAlaGlyValGluAlaVal 494

  859 TTGCAAACAGGC.........................AACCCCTATATAGG 884
        :::||||||                        |||        ||
  495 SerAsnThrGlyThrGlnLeuAsnThrAlaMetAlaAsnLeuGlnAsnGl 511

  885 AAAAAGCAATGGCTTCCAGCTAGTTCGTAAAGATTGGTTCTATGATGAAA 934
      |   ::::::      :::   :::   ::::::   ::::::|||
  511 yIleAsnAspLysAlaAsnThrLeuAlaSerGluAsnTyrHisAspAla. 527

  935 TCTTTGCTGGAGATACCCATTCAGTATTCTACGAACCACATCAAAATGGG 984
              :::|||      :::::::::   :::          ||:::
  528 .....AspSerAspLysLysThrAlaTyrThrGlnAlaValThrAsnAla 542

  985 AAATACTTTTTTAACGACAATAAT.......................... 1008
      :::   :::::::|||   |||:::
  543 GluAsnIleLeuAsnLysAsnSerGlySerAsnLeuAspLysAlaAlaVa 559

 1009 .........................AATGGCGCAGGAAAAATCGATGCCA 1033
                                |||:::   |||   ::::::::::
  559 lGluAsnAlaLeuSerGlnValThrAsnAlaLysGlyAlaLeuAsnGlyA 576

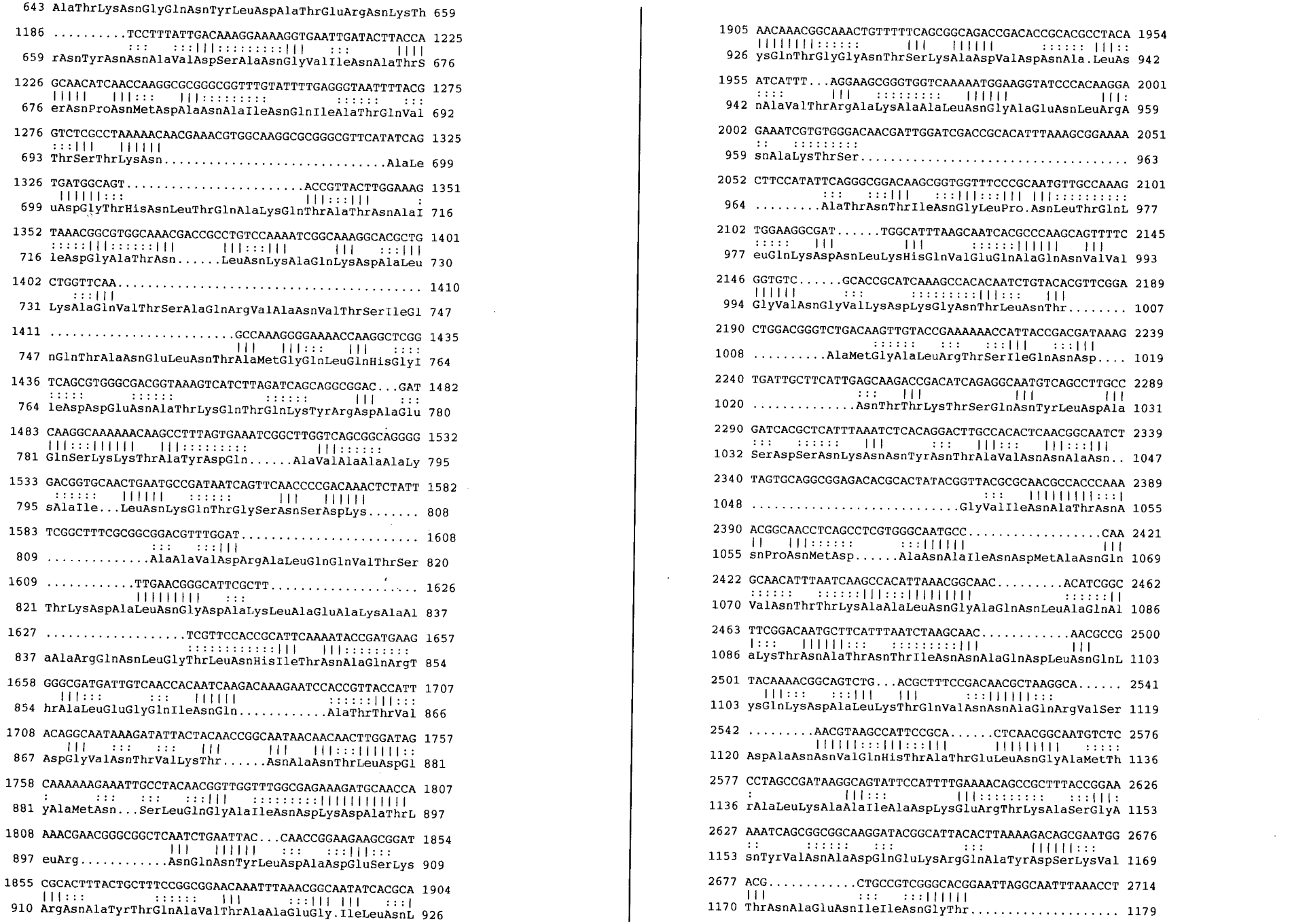
 1034 AACAT.......................................... 1038
      ::|||
  576 snHisAsnLeuGluGlnAlaLysSerAsnAlaAsnThrThrIleAsnGly 592

 1039 ...AAACACTATTCTCTACCTTAT......AGATTAAAAACACGAACCGT 1079
         :::|||   :::              :::|||||||   ::::::
  593 LeuGlnHisLeuThrThrAlaGlnLysAspLysLeuLysGlnGlnValGl 609

 1080 TCAATTGTTTAATGTTTCTTTATCCGAGACAGCAAGAGAACCTGTTTATC 1129
       |||      ||||||:::      :::|||:::::::::   :::
  609 nGlnAlaGlnAsnValAlaGlyValAspThrValLysSerSerAlaAsnT 626

 1130 ATGCTGCAGGTGGGGTCAACAGTTATCGACCCAGACTG............ 1167
             |||:::::::::::::   |||      :::
  626 hrLeuAsnGlyAlaMetGlyThrLeuArgAsnSerIleGlnAspAsnThr 642

 1168 ......AATAATGGAGAAAATATT....................... 1185
            :::|||||||:::|||
```





```
1700 lThrAsnAla.....................LysThrGlnLeuAsnGlyA 1710

4282 ..............AGTGCGGAATGGGGCGTAAACGCCGAAATCAAAGGT 4317
                    |||:::   ::::::|||:::   |||   |||
1710 snHisAsnLeuGluValAlaLysGlnAsnAlaAsnThrAlaIleAspGly 1726

4318 TTCACGCTGTCCCTCCACGCTGCCGCCGCCAAGGGGCCGCAA 4359
     :::|||                 :::   :::|||||||||
1727 LeuThr..................SerLeuAsnGlyProGln 1734
```